Genome Editing for Low-Acrylamide Wheat (*Triticum aestivum*)

Sarah Raffan

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Science

School of Biological Sciences

December 2020

Word count: 59,690

Abstract

Acrylamide (C_3H_5NO) is a food processing contaminant that has been classed as a probable (Group 2a) human carcinogen. Acrylamide forms from the reaction of free (non-protein) asparagine with reducing sugars during food processing. All major cereal products are affected and wheat products represent one of the main sources of dietary acrylamide intake in Europe.

Asparagine concentration is the determining factor for acrylamide formation in cereal products. Asparagine biosynthesis is catalysed by a family of enzymes called asparagine synthetases (ASNs). The *ASN* genes were investigated and five *ASN* genes (*TaASN1-4*, with a double copy of *TaASN3*) identified in wheat (*Triticum aestivum*), with *TaASN2* showing grain-specific expression.

CRISPR/Cas9 was used to knock out the *TaASN2* gene of wheat cv. Cadenza. A polycistronic gene containing four gRNAs, interspaced with tRNAs, was designed and introduced into wheat embryos by particle bombardment. The subsequent edits were characterised in the T1 and T2 generations using Next Generation Sequencing nucleotide sequence analysis. Triple (A, B, and D genome) nulls were identified, alongside an AD and an A genome null.

Amino acid concentrations were measured in the T2 and T3 seed, with one triple null line showing a substantial reduction in the free asparagine concentration in the grain (90 % in the T2 seed and 50 % in the T3 seed compared with wildtype). The free asparagine also reduced as a proportion of the total free amino acid pool. Significant effects were also seen in glutamate and aspartate concentrations. Free asparagine and total free amino acid concentrations were higher in the T3 than T2 seeds, probably due to heat stress, but the concentrations in the edited plants remained substantially lower than in wildtype. Some of the edited lines showed poor germination, but this could be overcome by application of exogenous asparagine and no other phenotype was noted.

Acknowledgements

First and foremostly, I would like to thank my supervisor, Professor Nigel Halford, for his advice and assistance throughout this doctorate. His solid guidance and patience at my stubbornness has underpinned this project. I would further like to thank my academic supervisor Prof. Keith Edwards for the initial work on the project and for his continued support and input.

Special thanks must go to Dr Damiano Martignago for his help and guidance during the early stages of this project. I also have to thank the GM Glasshouse team at Rothamsted for the time and care they put into looking after my plants: particular thanks go to Helen Martin for steadfast determination to get the T2 seeds to germinate and the love with which she took care of them. Furthermore, I owe the entire Cereal Transformation team at Rothamsted thanks for their support and for how warm and welcoming they were, particularly Dr Lucy Hyde for performing the biolistic bombardments, and Melloney St-Leger for her work on the T0 generation. The Bristol Genomics facility provided the NGS sequencing and Paul Wilkinson deserves so much thanks for all the time and effort he put into teaching me how to analyse it.

I would like to thank KWS UK Ltd, and Dr Ed Byrne in particular, for my three months of invaluable field experience, as well as insights into the plant breeding industry. The whole team was so welcoming and, although the work was occasionally tough, I can now confidently tell the difference between wheat and barley.

Finally, this would not have been possible without the support of my family and friends over the entire doctorate. The emotional support, patience at being continually lectured about acrylamide and endless proofreading have been critical in this journey!

This work was funded by an iCASE studentship from the South West Biosciences Doctoral Training Partnership (SWBio DTP), with partners: Rothamsted Research, University of Bristol, Agriculture and Horticulture Development Board, KWS UK Ltd, Saaten Union UK Ltd, RAGT Seeds Ltd, Syngenta UK Ltd, and Limagrain UK Ltd.

Author’s Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED:  DATE: 15/12/2020

Table of Contents

List of Figures

List of Tables

Chapter 7: *TaASN2*-knockout Phenotype and Amino Acid Analysis

Abbreviations

Units of Length

m	metre
cm	centimetre (10^{-2} metre)
mm	millimetre (10^{-3} metre)

Units of Volume

L	litre
mL	millilitre (10^{-3} litre)
µL	microlitre (10^{-6} litre)

Units of Mass

kg	kilogram (10^{3} g)
g	gram
mg	milligram (10^{-3} g)
µg	microgram (10^{-6} g)
ng	nanogram (10^{-9} g)

Units of Concentration

M	molar
mM	millimolar
mg/mL	milligram per millilitre
µg/µL	microgram per microlitre
µg/kg	milligram per kilogram
mmol/kg	millimole per kilogram
pmol/µL	picomol per microlitre
g/L	gram per litre
ppb	parts per billion

Other Units

bp	nucleotide base pair
kbp	kilobase pair (1000 bp)
FPKM	fragments per kilobase mapped

°C	degree Celcius (temperature)
rpm	revolutions per minute
kDa	kilodalton (molecular weight)

Nucleic Acid Abbreviations

A	adenine
C	cytosine
G	guanine
T	thymine
H	adenine / cytosine / thymine

Amino Acid Abbreviations

Ala/A	alanine
Arg/R	arginine
Asn/N	asparagine
Asp/D	aspartic acid (aspartate)
Cys/C	cysteine
Glu/E	glutamic acid (glutamate)
Gln/Q	glutamine
Gly/G	glycine
His/H	histidine
Ile/I	isoleucine
Leu/L	leucine
Lys/K	lysine
Met/M	methionine
Phe/F	phenylalanine
Pro/P	proline
Ser/S	serine
Thr/T	threonine
Trp/W	tryptophan
Tyr/Y	tyrosine
Val/V	valine

Materials and chemicals

AMP	adenosine monophosphate
ATP	adenosine triphosphate
EDTA	ethylenediamine tetraacetic acid
Mg^{2+}	magnesium ion
$MgCl_2$	magnesium chloride
NaCl	sodium chloride
NADH	nicotinamide adenine dinucleotide + hydrogen
NH_3	ammonia
SO_3	sulphur trioxide
SOC	Super optimal broth with catabolite repression
Tris	tris(hydroxymethyl)aminomethane

Other abbreviations

ASN	asparagine synthetase
BLAST	Basic Local Alignment Search Tool
cDNA	complementary DNA
CEV	N-(2-carbamoylethyl)valine
CRISPR	clustered regularly interspaced short palindromic repeat
Cas9	CRISPR-associated protein 9
cv.	cultivar
DNA	deoxyribonucleoic acid
DSB	double-stranded break
dsDNA	double stranded DNA
EMS	ethyl methanesulfonate
gRNA	guide RNA
HDR	homology-directed repair
MOE	margin of exposure
mRNA	messenger RNA
NHEJ	non-homologous end joining
PAGE	polyacrylamide gel electrophoresis
PAM	protospacer adjacent motif
PCR	polymerase chain reaction
PTG	polycistronic tRNA-gRNA

QTL	quantitative trait locus or loci
RNA	ribonucleic acid
RNAi	RNA interference
SNP	single nucleotide polymorphism
SSN	sequence-specific nucleases
TALEN	transcription activator-like effector nuclease
T-DNA	transfer DNA
TILLING	Targeting Induced Local Lesions in Genomes
tRNA	transfer RNA
ZFN	zinc finger nuclease

Organisations

BBSRC	Biotechnology and Biological Sciences Research Council, UK
CONTAM	European Food Safety Authority Panel on Contaminants in the Food Chain
EFSA	European Food Safety Authority
EU	European Union
FAO	Food and Agriculture Organisation
IARC	International Agency for Research on Cancer
IWGSC	International Wheat Genome Sequencing Consortium
JECFA	Joint Expert Committee on Food Additives
NCBI	National Centre for Biotechnology Information
WHO	World Health Organisation

Chapter 1

Introduction and Literature Review

Foreword

The information in this review has been published in:

Raffan, S. and Halford, N.G., 2019. Acrylamide in food: Progress in and prospects for genetic and agronomic solutions. *Annals of Applied Biology*, *175*(3), pp.259-281.

1.1. Acrylamide

Acrylamide (C_3H_5NO; Figure 1.1) is an organic compound which is classified as a serious health hazard with acute toxicity in the European Union. It is white, odourless, crystalline and water-soluble. Acrylamide forms a non-toxic polymer, polyacrylamide (Figure 1.1), which has widespread industrial uses including being an important component of polyacrylamide gel electrophoresis (PAGE), a technique to separate biological macromolecules, usually proteins or nucleic acids. Although polyacrylamide is non-toxic, it may contain a small concentration of monomeric acrylamide as an impurity. Since acrylamide, the monomer, is considered toxic, and polyacrylamide is used as a flocculant in wastewater and sewage treatment, acrylamide is a potential water pollutant. The World Health Organisation has set a guideline acceptable limit of 0.5 µg per litre, or parts per billion (µg/kg).

Acrylamide has neurotoxic and developmental effects, as well as affecting male reproduction, and is known to be carcinogenic in laboratory animals (reviewed by Tyl and Friedman, 2003; CONTAM Panel (European Food Safety Authority Panel on Contaminants in the Food Chain), 2015). As a result, it has been classed as a Group 2A carcinogen (meaning that it is probably carcinogenic to humans) by the International Agency for Research on Cancer (IARC) (International Agency for Research on Cancer, 1994), the specialised cancer agency of the World Health Organization. Acrylamide can be absorbed through the skin, inhaled or ingested, and when ingested it can readily be absorbed through the gastrointestinal tract, from where it can spread to all tissues in the body. Once in the body, it can be metabolised to glycidamide ($C_3H_5NO_2$, Figure 1.1), a highly reactive electrophilic epoxide, through oxidation by a cytochrome P450 (Segerbäck *et al*., 1995; Ghanayem *et al*., 2005; Sumner *et al*., 1999). Glycidamide may be responsible for some or all of the genotoxic and carcinogenic effects attributed to acrylamide (reviewed by CONTAM Panel, 2015; Zhivagui *et al*., 2019; Von Tungeln *et al*., 2012; Besaratinia and Pfeifer, 2003; 2004).

Acrylamide is found in tobacco smoke, with one of the key sources of information on the toxicity of acrylamide to humans coming from studies on smokers. The other main source is from occupational exposure, due to polyacrylamide's industrial uses. Most of the common markers for acrylamide exposure arise from reactions of acrylamide and glycidamide with proteins, notably haemoglobin in the blood where acrylamide forms the adduct N-(2-carbamoylethyl)valine (CEV) from reaction with valine residues located at the N-termini of globin chains. Acrylamide exposure is measured by determining the ratio of adducts to globin chains, with Bergmark *et al*. (1993) showing that high levels of these adducts were detectable in factory workers manufacturing acrylamide, and that levels of adducts correlated with peripheral neuropathy. It was then shown that the levels of CEV adducts were higher in laboratory workers using PAGE gels than in non-smoking controls (Bergmark, 1997); however, the levels in the non-smoking controls in that study were higher than was expected. A

possible cause of this was identified when CEV adducts were discovered in rats fed fried animal feed, with acrylamide forming in the frying process (Tareke *et al*., 2000). It was postulated that cooked food could be a major source of acrylamide exposure, and this was confirmed in 2002 when acrylamide was shown to be present in many common cooked foods (Tareke *et al*., 2002). The formation of acrylamide was connected to frying, baking, toasting or roasting processes (low moisture, high temperature cooking environments), with no acrylamide being detected in the boiled foods analysed (Tareke *et al*., 2002). As no acrylamide was detected in the raw food stuff, acrylamide was classified as a food processing contaminant (Curtis *et al*., 2014b), which means it is a substance that is produced during the cooking or processing of foodstuffs and is unwanted either because it negatively affects product quality or, as in this case, is potentially harmful.

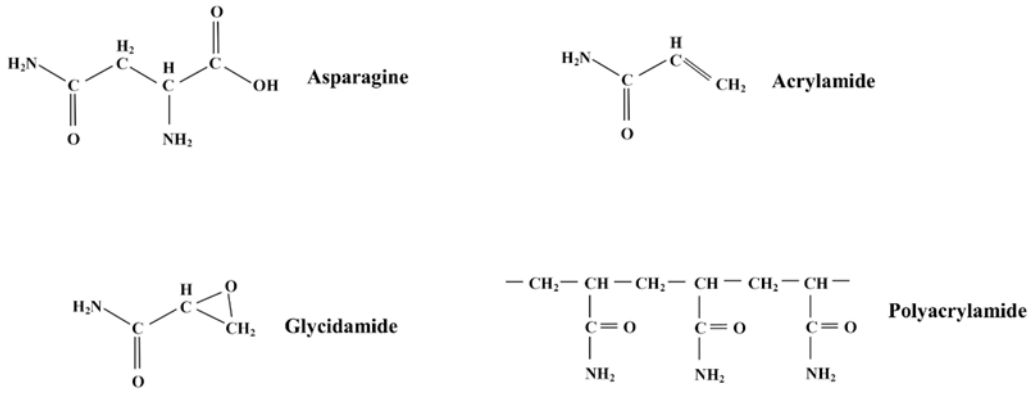


Figure 1.1. The structures of (clockwise from top left) asparagine ($C_4H_8N_2O_3$ precursor to acrylamide), acrylamide (C_3H_5NO), polyacrylamide (polymer of acrylamide), and glycidamide ($C_3H_5NO_2$, metabolite of acrylamide). Adapted from Raffan and Halford, 2019.

1.2. Risk represented by dietary acrylamide intake

After the discovery of acrylamide in food, an opinion on the risks it posed was published by the United Nations' Food and Agriculture Organisation (FAO) and World Health Organisation (WHO) Joint Expert Committee on Food Additives (JECFA). This opinion (JECFA, 2006) stated that dietary exposure alone did not pose a risk for adverse neurological effects, although high levels of exposure could pose a risk for neural morphological changes; however, when considering the genotoxic and carcinogenic effects, the margins of exposure (MOE) indicated a health concern. This opinion on acrylamide was restated by JECFA in 2011 (JECFA, 2011), and in the European Food Safety Authority (EFSA) Panel on Contaminants in the Food Chain (CONTAM Panel) report in 2015 (CONTAM Panel, 2015).

A MOE is used in risk assessments to assess how dangerous genotoxic and carcinogenic substances are. It is a ratio of the no-observed-adverse-effect level to the estimated exposure dose. The mean estimated dietary acrylamide exposure was calculated by JECFA to be 1 µg per kg bodyweight per

day, with the highest consumers of acrylamide exposed to 4 µg per kg bodyweight per day. The CONTAM panel estimated that the mean exposure ranged from 0.4 to 1.9 µg per kg body weight per day, with the 95th percentile ranging from 0.6 to 3.4 µg per kg body weight per day (CONTAM Panel, 2015). These estimates are far lower than the amounts used in the toxicology studies carried out on rodents (Beland *et al*., 2013; 2015; Tyl and Friedman, 2003; Sumner *et al*., 2003), which raises questions on the applicability of the rodent data to human risk levels. Whilst it is undisputed that acrylamide causes cancer in rodents (Beland *et al*., 2013; 2015; CONTAM Panel, 2015), a link between human cancers and acrylamide exposure has not been conclusively demonstrated. Rodent toxicology studies are based on high doses over a short period of time, rather than low doses over a long period of time as is seen in human dietary exposure, and the data would also have to be extrapolated from a rat to a human system. Even studies into occupational exposure involve levels of acrylamide that are much higher than those estimated from dietary exposure (Exon, 2006).

The epidemiological studies that have been conducted to assess the carcinogenic potential of acrylamide in humans have been limited and like all epidemiological studies may be confounded by other variables. The potential role of dietary acrylamide in renal (Hogervorst *et al*., 2008), endometrial and ovarian cancers (Hogervorst *et al*., 2007; Pelucchi *et al*., 2015) has been explored; however, studies based on questionnaires about diet have obvious limitations, while methods using direct detection (such as the detection of haemoglobin-adduct levels) might not give an assessment of past acrylamide exposure. The CONTAM report (2015) concluded that the epidemiological studies were not consistent enough to provide solid evidence on which to draw conclusions on the risk of dietary acrylamide intake. Instead, the panel based its opinion on the rodent toxicology studies, its assessment reinforcing the opinion of the JECFA, and leading to the imposition of stronger risk management measures from the European Commission (Section 1.5).

Zhivagui *et al*. (2019) identified a unique mutational signature imprinted by glycidamide and showed that this mutational signature was present in genomes of human tumour cells. The signature was found in around one-third of approximately 1600 tumour genomes investigated. Since glycidamide is found in cigarette smoke, other mutational signatures, such as from benzo[a]pyrene, the major mutagen in tobacco smoke, were considered to distinguish the mutagenesis arising from dietary and smoking-related acrylamide exposure. A subset of 184 liver tumour samples and 217 tumours of 15 other cancer types were found to carry the glycidamide mutational signature, but not that of benzo[a]pyrene, indicating that these mutations were the result of dietary or occupational acrylamide exposure. No evidence was given that mutations present in the tumour genomes were responsible for the formation of those tumours; nevertheless, this represents the strongest evidence yet linking dietary acrylamide exposure with cancer risk. The true danger of dietary acrylamide is still debated, with Eisenbrand (2020) claiming that the risk of dietary acrylamide intake is overestimated, with the genotoxicity of acrylamide occurring at exceedingly high dose levels, if at all.

1.3. Levels of acrylamide in popular foods

Since the discovery of acrylamide in cooked foodstuffs (Tareke *et al*., 2002), the levels of acrylamide in different foods have been monitored, with data first collected in 2003 by the European Commission's Joint Research Centre's Institute for Reference Materials and Measurements. After 2007, Member States collected the data on acrylamide levels and supplied it to EFSA. EFSA analyses the acrylamide data and publishes reports which inform the Commission's risk management measures (EFSA, 2009; 2010; 2011; 2012; CONTAM Panel, 2015).

The 2015 CONTAM report also contained data from six food associations, with the acrylamide data shown in Figure 1.2. Vegetable crisps and coffee substitutes showed the highest concentrations of acrylamide, with chicory-based coffee substitute showing a mean concentration of 2942 µg/kg. Dry coffee only showed a mean of 522 µg/kg; however, it is consumed much more frequently. Mean values in the hundreds of µg/kg were seen for potato crisps, chips (fries), biscuits, breakfast cereals and crispbreads, which is much higher than the tolerance level for acrylamide in drinking water (0.5 µg/kg). Estimates on the contributions of different food groups to dietary acrylamide intake in the different member states were calculated (EFSA, 2011; CONTAM Panel, 2015; Raffan and Halford, 2019; Figure 1.2), and showed that food groups which were high in acrylamide, such as vegetable crisps and substitute coffee, were not the biggest contributors to dietary acrylamide intake. Indeed, bread was the second highest contributor to dietary acrylamide intake, despite being relatively low in acrylamide, as it is consumed so regularly and is often toasted, which increases the acrylamide concentration from tens of µg/kg to several hundred.

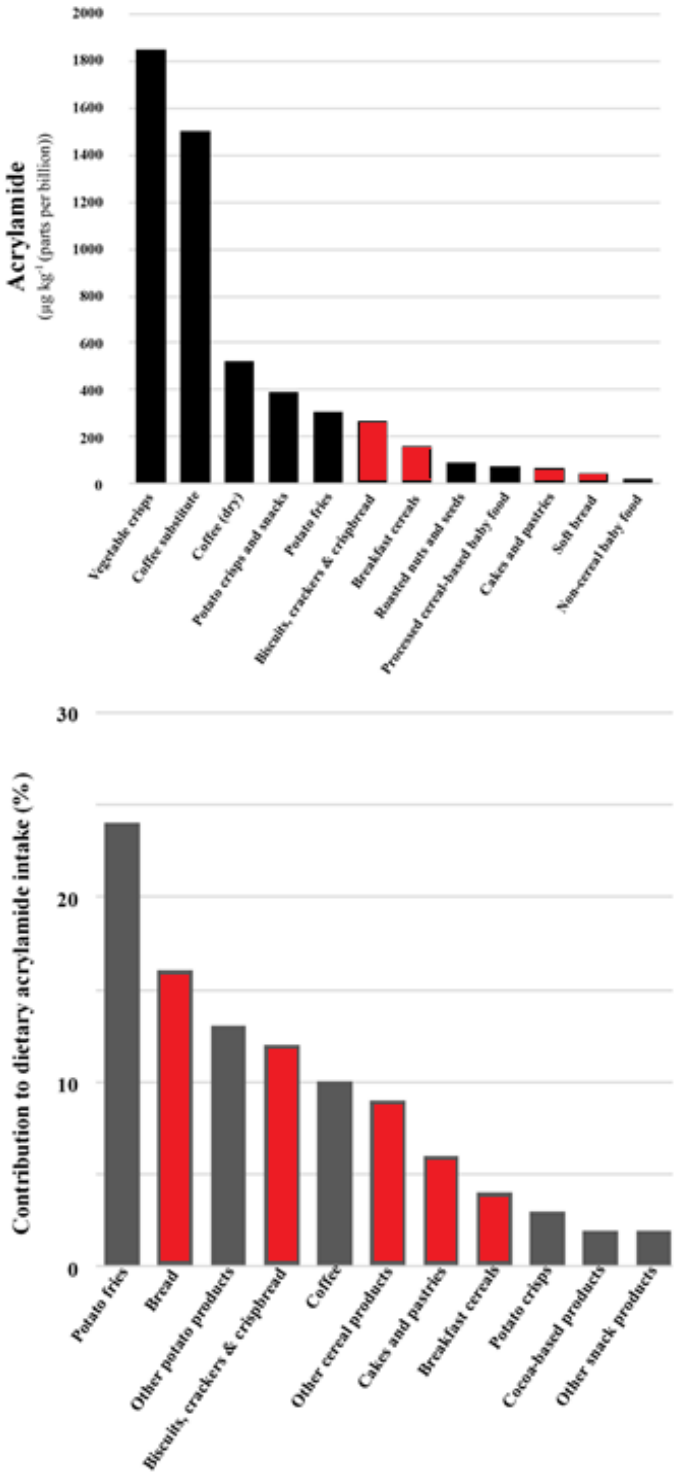


Figure 1.2. Graphs showing: Top: Mean acrylamide levels in different food groups. Bottom: The contributions of different food groups to dietary acrylamide intake in the European Union. Categories containing wheat products are highlighted in red. Taken from Raffan and Halford (2019) and based on data provided in the EFSA Panel on Contaminants in the Food Chain report of 2015 (CONTAM Panel, 2015).

1.4. Acrylamide formation in foodstuffs

Acrylamide forms in the Maillard reaction from free (soluble, non-protein) asparagine (Figure 1.1) and reducing sugars (Mottram *et al*., 2002; Stadler *et al*., 2002). The Maillard reaction is named after French chemist Louis Camille Maillard, who first described it, in French, in 1912 (Maillard, 1912); however, the reaction is now understood to follow the steps laid out in the Hodge Scheme (Hodge, 1953).

The Maillard reaction (reviewed in Martins *et al*., 2000) is actually a complex series of non-enzymatic reactions that take place between reducing sugars, such as glucose, fructose and maltose, and free amino acids under high-temperature (>120 °C), low-moisture conditions, such as occur during frying, roasting, baking and toasting. It is initiated by the generation of a Schiff base (a compound carrying an imine or azomethine functional group) from the condensation of an amino group from an amino acid with the carbonyl group of a reducing sugar. Amadori or Heyns rearrangement products are then produced from the cyclisation and acid-catalysed rearrangement of the Schiff base, depending on the reducing sugar involved, and these go on to form various sugar dehydration and fragmentation products, from fragmentation, deamination, enolisation and dehydration reactions. These sugar dehydration and fragmentation products contain one or more carbonyl groups and are highly reactive. They go on to participate in further reactions with other amino acids, such as a Strecker degradation, where an amino acid undergoes deamination and decarboxylation to form an aldehyde, an α-aminoketone and carbon dioxide.

Acrylamide results from the Strecker-type degradation of asparagine (Zyzak *et al*., 2003). This occurs through the generation of a decarboxylated Schiff base which then decomposes directly to acrylamide, or through the elimination of a carbonyl group to produce an intermediate, 3-aminopropionamide. 3-aminopropionamide is then converted to acrylamide by the removal of ammonia (Granvogl and Schieberle, 2006; Granvogl *et al*., 2007; Hedegaard *et al*., 2008). Thus, both free asparagine and reducing sugars can be considered to be the precursors for acrylamide formation in foodstuffs.

The Maillard reaction products are responsible for the colours (melanoidin pigments), flavours and aromas that are associated with roasted, fried, toasted and baked foods (Mottram, 2007; Halford *et al*., 2011). As the different Maillard reaction products are responsible for different flavours and colours, measures to reduce acrylamide might also affect the defining characteristics of specific products. For instance, there is a strong relationship between colour and acrylamide (Halford *et al*., 2012), because the colours in cooked foods form *via* similar pathways to acrylamide in the Maillard reaction. Colour can therefore be an indicator of acrylamide concentration, and this was the basis for the 2017 ‘Go for Gold’ campaign launched by the UK Food Standards Agency (Food Standards Agency, 2017). Colour

has also become an important quality control parameter as manufacturers seek to limit the formation of acrylamide in their products.

Acrylamide has been detected in dried fruit, such as dates and prunes (De Paola *et al*., 2017; Bermudo *et al*., 2006), even though high temperatures are not used in the drying process and so the Maillard reaction cannot be occurring. An alternative source for acrylamide formation has also been proposed in wheat products through pyrolytic formation in heated gluten (Claus *et al*., 2006b; Weißhaar *et al*., 2004; Amrein *et al*., 2004). Acrylamide concentration in the cooked flour has been shown to correlate well with asparagine concentration in the raw flour (Curtis *et al*., 2014a), indicating that these other pathways for acrylamide formation play a minor role, if any role at all, for acrylamide formation in cereal products.

1.5. Regulations on acrylamide levels in food

The publication of the Tareke (2002) study led to widespread concern in the food industry. As a result, the industry has developed a variety of methods to reduce acrylamide levels and these have been compiled by FoodDrinkEurope into an 'Acrylamide Toolbox' (FoodDrinkEurope, 2019). Meanwhile, EU Member States have been required to collect data on the acrylamide levels in food, and these data are reported to EFSA so that assessments can be made of the success of the mitigation methods in reducing acrylamide in food. The European Union's regulations on acrylamide have been guided by reports made by EFSA based on these data. Recommendation 2007/331/EC was issued in 2007 to require the continued monitoring of acrylamide in food up to 2009, with EFSA reporting in 2009 that no general lowering of acrylamide levels across food groups had been achieved (EFSA, 2009). Consequently, Recommendation 2010/307/EU (European Commission, 2010) required Member States to continue collecting acrylamide data.

A second EFSA report (EFSA, 2010) declared that there was a trend towards low acrylamide levels in certain food groups, but that other groups showed a rise in acrylamide levels. Commission Regulation C (2010) 9681(European Commission, 2011) followed, introducing Indicative Values for acrylamide in food (shown in Table 1.1). If a food exceeded its Indicative Value, the relevant authorities in the Member State were required to investigate it and ensure that the manufacturer took appropriate measures. These Indicative Values were not intended to be safety thresholds, but instead were set at levels that the European Commission believed were achievable for each food type, based on the monitoring data from EFSA.

Table 1.1. Indicative Values and Benchmark Levels for acrylamide in food (from Raffan and Halford, 2019), as set by the European Commission. Indicative Values 2011 come from Regulation C (2010) 9681, Indicative Values 2013 come from Recommendation 2013/647/EU and Benchmark Levels 2017 come from the Regulation (EU) 2017/2158.

Food	Indicative Value 2011 (ppb)	Indicative Value 2013 (ppb)	Benchmark Level 2017 (ppb)
Baby foods (not cereal based) without prunes	80	50	40
Baby foods (not cereal based) with prunes	80	80	40
Cereal-based baby foods	100	50	40
Soft bread (wheat)	150	80	50
Soft bread (other)	150	150	100
Biscuits and rusks for infants and young children	250	200	150
Breakfast cereals: maize, oat, spelt, barley and rice based	400	200	150
Breakfast cereals: wheat and rye based	400	300	300
Breakfast cereals: bran products, whole grain cereals, gun puffed grain	400	400	300
Roast coffee	450	450	400
Crispbread	500	450	350
Biscuits	500	500	350
Crackers	500	500	400
French fries	600	600	500
Instant coffee	900	900	850
Potato crisps	1000	1000	750
Gingerbread	-	1000	800
Coffee substitute (cereal-based)	-	2000	500
Coffee substitute (chicory)	-	4000	4000

Subsequent reports from EFSA were issued in 2011 and 2012 (EFSA, 2011; 2012) and reported little improvement in acrylamide levels, with categories such as instant coffee and French fries showing

increased acrylamide levels. This led to Recommendation 2013/647/EU (European Commission, 2013) which revised the Indicative Values, reducing them in many cases (Table 1.1). An assessment of the risk posed by dietary acrylamide was then published in 2015 (CONTAM Panel, 2015), resulting in Commission Regulation (EU) 2017/2158 (European Commission, 2017), which represented further strengthening of the risk management regulations for acrylamide, and came into force in April 2018.

This Regulation replaced the previous Indicative Values with Benchmark Levels, lowering the values in almost all cases, although these Benchmark Levels were described as performance indicators. The reduction did not reflect any decrease seen in the monitoring data. Potato crisp data from this period, for example, did not show any reduction, although a reduction of over 50 % had been achieved from 2002 to 2011 (Powers *et al*., 2013; 2017). This might reflect that the easiest and most effective acrylamide-reduction methods had already been devised and implemented by 2011. The data also showed that consistent compliance with the regulations might be difficult to achieve, due to factors such as the seasonal effect (Powers *et al*., 2013; 2017). As European potatoes (*Solanum tuberosum*) are harvested in a particular window in the calendar, between July and October, potatoes required outside of this period come from stored stocks. Stored potatoes suffer from cold and senescent sweetening, which increase the levels of reducing sugars due to vacuolar invertase (VInv) activity (Clasen *et al*., 2016; Zhu *et al*., 2014; Wiberley-Bradford and Bethke, 2018). Senescent sweetening also results from the breakdown of starch, and this is catalysed by several enzymes, including phosphorylase L (PhL) and starch-associated R1 (R1) (Sowokinos, 2001). Another factor that affects regulatory compliance is a geographical one, with potatoes grown at a more northern latitude, such as in Denmark, Norway and Sweden, showing a higher acrylamide-forming potential than potatoes grown elsewhere (Powers *et al*., 2017).

Regulation (EU) 2017/2158 (European Commission, 2017) threatened the imposition of Maximum Levels, i.e. the setting of maximum acrylamide concentrations at which it would be legal to sell a food product. Mitigation measures to reduce acrylamide formation were laid out in the annexes of Regulation (EU) 2017/2158. These included food processing measures, as well as guidance on crop management and variety selection, and effectively formed codes of practice for the industry. Many of these measures came from the industry-devised Acrylamide Toolbox (FoodDrinkEurope, 2019). Businesses are also required to monitor the acrylamide levels in their products, with responsibility of ensuring compliance with the Regulation falling on Member States.

In the United States, there is nothing like the EU's Regulations on acrylamide, but the Food and Drug Administration (FDA) has issued guidance for the food industry (Food and Drug Administration, 2016). However, a lawsuit was filed by the Attorney General of the State of California against four fast-food chains (Burger King, KFC, McDonald's, and Wendy's) and five manufacturers of potato products (Frito-Lay, H.J. Heinz, Kettle Foods, Lance Inc and Procter & Gamble) in 2005. This lawsuit

sued the companies for not adding a Proposition 65 warning to their products to inform customers of the presence of acrylamide. Proposition 65 is California's 'Safe Drinking Water and Toxic Enforcement Act'. The lawsuit was settled in 2008, with the companies pledging to cut acrylamide levels in their products, paying $3 million in fines and posting warnings on their products or in their fast food restaurants. The Council for Education and Research on Toxics then sued Starbucks in 2010, alongside 90 other companies in the coffee industry, demanding that products be labelled, or that acrylamide be removed from the products (an impossible demand).

1.6. Reducing the acrylamide-forming potential of wheat

Free asparagine is the determining factor in acrylamide formation in products made from cereal grains (Muttucumaru *et al*., 2006; Granvogl *et al*., 2007; Curtis *et al*., 2009; 2010; 2016; Postles *et al*., 2013). Reducing free asparagine accumulation in the grain would, therefore, lower the acrylamide-forming potential of cereal grain. The free asparagine concentration in the grain is known to vary with wheat (*Triticum aestivum*) variety (Taeymans *et al*., 2004; Curtis *et al*., 2018b), and can also be affected by agronomic practice (Claus *et al*., 2006a; Curtis *et al*., 2016; 2018b; Martinek *et al*., 2009). The Cereals and Oilseeds department of the Agriculture and Horticulture Development Board (AHDB), formerly known as the Home Grown Cereals Authority, together with Rothamsted Research, compared six wheat varieties at six different field sites over two years and revealed large varietal differences, alongside site and year effects, and interactions between these factors (Curtis *et al*., 2009). The varieties Claire and Robigus were used in the study. These are both Group 3 wheat varieties (National Association of British and Irish Millers), commonly known as biscuit wheats and used for products such as biscuits, cakes and breakfast cereals. Of the two, Robigus showed a higher mean asparagine value, with a difference of 37 % between the varieties; however, Robigus also showed a much higher range of free asparagine concentrations. The role of variety in free asparagine accumulation was confirmed in a field trial of winter wheats (Curtis *et al*., 2018b), where concentrations ranged from 0.71 mmol/kg to 11.29 mmol/kg.

Whilst environmental factors can influence varietal effects, wheat varieties that show consistently low asparagine levels have been identified (Curtis *et al*., 2018b). Five out of the eight 'low' varieties were Group 3 varieties, two were Group 4 (feed wheat varieties) and one was Group 2 (varieties with bread-making potential but not suitable for all purposes); however, no significant effect of Group was seen on the asparagine concentration, with high asparagine varieties in all of the groups.

Rye (*Secale cereale*) has also shown environmental effects and varietal differences in free asparagine concentration (Curtis *et al*., 2010), with free asparagine concentration linked to bran yield. In another study, a UK field trial of five commercial rye varieties showed a nearly 50 % difference in the free asparagine concentration between the highest and lowest varieties (Postles *et al*., 2013). This

highlights that switching the variety used in a product may reduce the amount of acrylamide that forms.

Žilić *et al*. (2017) looked at the asparagine accumulation in eight cereal species (bread wheat, *Triticum aestivum*; pasta wheat, *Triticum durum*; spelt wheat, *Triticum spelta*; emmer wheat, *Triticum dicoccum*; rye, *Secale cereale*; barley, *Hordeum vulgare*; oats, *Avena sativa*; and maize, *Zea mays*). Rye showed the highest mean free asparagine concentration (8.93 mmol/kg), and bread wheat the lowest (3.23 mmol/kg). Varietal differences were seen within each species, with barley showing the largest range in free asparagine concentration (0.91 to 6.26 mmol/kg). However, it should be noted that this study examined a small number of varieties for each species.

1.7. Environmental effects and crop management for low asparagine cereal grains

Environmental factors, such as the nutritional status, particularly sulphur sufficiency (Muttucumaru *et al*., 2006; Granvogl *et al*., 2007), and pathogen infection (Curtis *et al*., 2016), also affect asparagine metabolism (Lea *et al*., 2007). These can be directly influenced, and even controlled, by crop management strategies and agronomy.

Nitrogen supply affects free asparagine accumulation in the grain (Winkler and Schön, 1980), with elevated nitrogen levels resulting in greater asparagine concentrations. This means that increasing nitrogen fertilization leads to higher free asparagine levels and more acrylamide formation in bread (Claus *et al*., 2006a; Martinek *et al*., 2009). Sulphur has the opposite effect (Shewry *et al*., 1983), with free asparagine accumulating under sulphur deficiency. In fact, under severe sulphur deficiency, up to 30-fold increases were seen in the free asparagine concentration in the grain of three varieties of wheat (Muttucumaru *et al*., 2006), with the total free amino acid concentration also rising. Nitrogen supply also leads to increased free asparagine accumulation in rye grain (Postles *et al*., 2013; 2016).

The effect of sulphur deficiency on wheat grain has been further confirmed through subsequent studies (Granvogl *et al*., 2007; Curtis *et al*., 2009; 2014a; 2018b). In the largest of these studies, wheat varieties in different milling groups responded differently to sulphur deprivation (Curtis *et al*., 2018b), with soft wheats in Group 4 most affected; however, there was also a variety effect, with the individual varieties within the groups showing differential responses as well. Varieties that showed low free asparagine accumulation under sulphur-sufficient conditions often showed high free asparagine concentration under sulphur starvation, meaning that the varietal ranking for asparagine accumulation was not consistent in the face of environmental effects. The exceptions amongst the field trials analysed in these studies were two field trials conducted in the 2011-2012 season. These showed no relationship between sulphur deficiency and free asparagine concentration (Curtis *et al*., 2014a; 2018b). One reason given for this was the potential interaction of sulphur feeding and

environmental effects as the season showed particularly unusual weather.

As a result of these studies, farmers have been advised to apply 20 kg sulphur (50 kg SO_3 equivalent) per hectare to wheat (AHDB, 2014), with the need to 'maintain balanced sulphur levels in the soil and to ensure correct nitrogen application' written into Regulation (EU) 2017/2158 (European Commission, 2017).

Despite showing a similar response to nitrogen, there is no observable effect of sulphur deprivation on free asparagine accumulation in rye, at least in the field (Postles *et al*., 2013). It was postulated that rye may be better able to obtain sulphur from the soil, or that it responds differently to sulphur starvation than wheat. Wheat may use free asparagine as a nitrogen store, accumulating it when there is inadequate sulphur to make sulphur-rich proteins (Zhao *et al*., 1999); however, this response is not seen in rye (Postles *et al*., 2013).

Biotic stress has also been shown to increase asparagine accumulation in many plant species (Lea *et al*., 2007). A two-year study in the Czech Republic suggested that effective fungicide control could lower asparagine concentration (Martinek *et al*., 2009), although this effect was only seen in one of the two growing seasons in the study. A larger UK trial using 47 wheat varieties showed an increase in asparagine concentration in all varieties when fungicide treatment was withheld (Curtis *et al*., 2016), although, due to severe disease, only 24 out of the 47 varieties could be harvested. Varietal differences were seen in the asparagine response to the fungal pathogens, with the increase in asparagine concentration ranging from 17 % (cv. Epson) to 135 % (cv. Cleveland), meaning that identifying 'low' varieties was difficult as the varietal ranking changed under biotic stress. However, unlike with sulphur deficiency, free asparagine was not the predominant amino acid in the grain (Curtis *et al*., 2016). Aspartic acid was the most abundant free amino acid in both the fungicide-treated and untreated conditions. Free asparagine also accumulates in response to a mycotoxin, deoxynivalenol, produced by *Fusarium graminearum*, the fungus that causes Fusarium head blight (Warth *et al*., 2015). In a study in Nebraska, asparagine concentrations were associated with disease pressure, as well as large kernel size and delayed harvest (Navrotskyi *et al*., 2018). Thus, disease control and mitigating biotic stress is very important in controlling asparagine accumulation, and this has been written into European Commission Regulation (EU) 2017/2158.

1.8. Genetic control of free asparagine accumulation in cereal grain

Asparagine plays an important role in nitrogen metabolism in plants and is a nitrogen transport and storage molecule in many species (Lea *et al*., 2007). Asparagine metabolism has been described in detail in a network generated by Curtis *et al*. (2018a) that includes 212 genes, enzymes or molecules and 246 reactions between them. Asparagine metabolism is underpinned by the core enzymes asparagine synthetase and asparaginase. Glutamine synthetase, NADH-dependent glutamate synthase,

ferredoxin-dependent glutamate synthase, glutamate dehydrogenase, aspartate amino transferase, glutamate decarboxylase, and aspartate kinase are also all key enzymes in the network.

Asparagine is generated from the ATP-dependent transfer of an amino group from glutamine to aspartate (Gaufichon *et al*., 2010; Xu *et al*., 2018). The reaction is catalysed by the enzyme asparagine synthetase (ASN) and produces asparagine and glutamate. The first *ASN* cDNAs were isolated from pea (*Pisum sativum*) (Tsai and Coruzzi, 1990) and two *ASNs* were identified, which encoded the enzymes AS1 and AS2. It is now known that there are three *ASN* genes in Arabidopsis (*Arabidopsis thaliana*), *AtASN1-3* (Lam *et al*., 1994; 1998). These genes are differentially expressed and regulated by different stimuli, including light, sucrose, stress and the supply of organic nitrogen (Lam *et al*., 1998). Potato has two *ASN* genes, *StASN1* and *StASN2* (Rommens *et al*., 2008). Wheat, barley and maize possess *ASN* gene families, with different *ASN* genes active in different tissues (Avila-Ospina *et al*., 2015; Duff *et al*., 2011; Gao *et al*., 2016; Todd *et al*., 2008).

In Arabidopsis, *AtASN1* expression is regulated by sucrose, being repressed by sucrose feeding in tissue cultures (Lam *et al*., 1998). Sucrose nonfermenting-1 (SNF1)-related protein kinase-1 (SnRK1) is a protein kinase in the plant sugar-sensing signalling pathway (reviewed by Hey *et al*., 2010), and experiments with reporter gene constructs have shown *AtASN1* expression to increase with the overexpression of *SnRK1* (Baena-González *et al*., 2007; Baena-González and Sheen, 2008; Confraria *et al*., 2013). The pathway is also underpinned by bZIP transcription factors, with low sucrose inducing *AtASN* via *AtZIP11* and high sucrose inhibiting *AtASN* gene expression *via AtbZIP9, AtbZIP10, AtbZIP25* and *AtbZIP63* (Hummel *et al*., 2009).

In wheat, another protein kinase, general control nonderepressible-2 (GCN2), which phosphorylates the α subunit of translation initiation factor eIF2 (eIF2α), affects *TaASN1* expression (Byrne *et al*., 2012). In fungi, GCN2 is known to regulate the balance between free amino acids and proteins (reviewed by Hinnebusch, 1992) and is activated in response to low levels of free amino acids. In Arabidopsis, it is activated in response to herbicides that inhibit amino acid biosynthesis (Zhang *et al*., 2003; 2008), as well as in part of the response to multiple stresses, including wounding, UV light and pathogen infection (Lageix *et al*., 2008). Transgenic wheat overexpressing *TaGCN2* showed a reduction in total free amino acid concentration, and asparagine in particular, as well as reduced *TaASN1* expression (Byrne *et al*., 2012). These transgenic lines did not show sulphur-induced *TaASN1* expression like the wildtype plants (Byrne *et al*., 2012).

1.9. The asparagine synthetase (*TaASN*) genes in wheat

Bread wheat contains five *TaASN* genes, *TaASN1, TaASN2, TaASN3.1, TaASN3.2* and *TaASN4*. *TaASN1*, *TaASN2* and *TaASN4* are located on chromsomes 5, 3 and 4 respectively, while the two *TaASN3* genes are located on chromosome 1 (Xu *et al*., 2018). Not much is known about *TaASN4* as

it was only recently identified, and it has not been described in any detail. The other *TaASN* genes show differential expression (Gao *et al*., 2016).

The expression of *TaASN2* is confined to the embryo and the endosperm and it is the most highly expressed *TaASN* gene in those tissues during mid to late grain development. The most responsive gene to nitrogen and sulphur levels elsewhere in the plant is *TaASN1* (Gao *et al*., 2016), which has also been shown to respond to salt and osmotic stress (Wang *et al*., 2005).

Both TaASN1 and TaASN2, when expressed heterologously, produce asparagine and glutamate from aspartate and glutamine, with the reaction described in a continuous Petri net model by Xu *et al*. (2018).

The main components of this (representing the asparagine synthetase as ASNe) are:

1) ASNe + glutamine → ASNe-NH_3 + glutamate

2) ASNe-NH_3 + Mg^{2+} + aspartate + ATP → ASNe + asparagine + AMP

In the assays described by Xu *et al*. (2018), glutamate accumulated at a faster rate than asparagine, and when both reactions plateaued, the concentration of glutamate was twice that of asparagine. Gaufichon *et al*. (2010) suggested that the reactions did not occur simultaneously, and the Xu *et al*. (2018) study seemed to confirm this, indicating that the earlier stages of the reactions, which produce glutamate, occurred before, and proceeded faster, than the later stages, which produce asparagine. Thus, glutamate synthesis occurs independently of asparagine synthesis and glutamate is still synthesized when aspartate is not available to make asparagine (Xu *et al*., 2018).

1.10. Prospects for reducing acrylamide-forming potential by plant breeding

The existence of acrylamide in food products is a major problem facing the food industry, and strategies to lower acrylamide levels through modifications to processing, described in the Acrylamide Toolbox (FoodDrinkEurope, 2019), could have knock-on effects on many of the compounds that impart colour, flavour and aroma, leading to changes in product characteristics. However, the acrylamide regulations in the European Union could be moving towards the setting of Maximum Levels, and if these were set at the current Benchmark Levels (Regulation (EU) 2017/2158), the food industry would struggle to achieve regulatory compliance in many products. Agronomic and genetic approaches could, therefore, lead to huge savings for the food industry by reducing the acrylamide-forming potential of the raw foodstuffs. This would negate the need to modify manufacturing lines and avoid changes to product characteristics, or could even prevent the loss of some products altogether.

Environmental effects on free asparagine concentrations have complicated plant breeding strategies for low acrylamide wheat. The identification of QTL for low asparagine, for example, has been slow. Some QTL were identified in 2018 using genome-wide association mapping (Rapp *et al*. 2018); however, the strongest of these only accounted for 18 % of the genotypic variance in free asparagine concentration. This trait showed no correlations with seed quality measures like protein content, sedimentation volume or falling number.

In potato, where reducing sugars are the limiting factor in acrylamide-formation, due to the relatively high concentration of free asparagine in potato tubers, the enzymes that produce, accumulate and turn over these reducing sugars are clear targets for genetic interventions. Low-sugar varieties, which have low glucose and low fructose concentrations, have already been bred for crisping to enable manufacturers to produce crisps with a uniform light brown colour. A potential QTL was identified for free asparagine and reducing sugars (Shepherd *et al*., 2010; 2013), with another QTL identified for fry colour (Bradshaw *et al*., 2008).

Free asparagine concentration does become limiting for acrylamide formation in potatoes if it becomes low enough, and GM technology has already been used to generate low asparagine, low acrylamide potatoes (Innate® and Innate® Generation 2, from Simplot Company in Boise, Idaho). Initially, expression of both *StASN* genes in potato was reduced by RNA interference (RNAi) but when they were grown in a field, the plants produced small, cracked tubers (Rommens *et al*., 2008). However, the tuber-specific targeting of *StASN1*, again by RNAi, produced phenotypically normal plants, with normal yields but reduced free asparagine concentrations (Chawla *et al*., 2012), and this trait underpins the Innate® and Innate® Generation 2 potatoes (USDA-APHIS, 2013; 2014). The tuber-specific knockdown of *StASN1*, and the subsequent reduction in free asparagine in potatoes, shows that the free asparagine that accumulates is produced *in situ* (Karley *et al*., 2002; Muttucumaru *et al*., 2014). Innate® Generation 2, which also has reduced expression of genes related to starch breakdown (*PhL*, (starch phosphorylase L), and *R1* (starch-associated R1)), reduced bruising through the knockdown of *PPO5* (polyphenol oxidase), and reduced susceptibility to cold and senescent sweetening through the reduced expression of a vacuolar invertase gene (*VInv*), have an acrylamide-forming potential only 10 % that of conventional potatoes. The Innate® Generation 2 potatoes also show greater resistance to *Phytophthora infestans*, the pathogen that causes late blight disease, through the incorporation of *Rpi-vnt1.1*, a resistance gene from a wild potato species, *Solanum venturi*. Genome editing has also been applied to potatoes, with transcription activator-like effector nucleases (TALENs) used to target the *VInv* gene (Clasen *et al*., 2016); knock-out lines had undetectable levels of reducing sugars in the tubers and lowered acrylamide levels in crisps produced from them.

1.11. The development of mutant lines

The generation and characterisation of mutants, both natural and induced, is vital to the study of gene function and metabolic pathways, and ultimately it is an essential part of plant breeding and crop improvement. Characterisation of natural genetic variation has been a fundamental resource over the last few decades, and it is the backbone of conventional plant breeding. Many studies have also utilised physical, chemical, or biological (e.g., T-DNA/transposon insertion) mutagenesis to generate, identify and characterise mutants, or construct mutant libraries, such as a transposon-inserted library in Arabidopsis (Kuromori *et al*., 2006), a T-DNA insert library in rice (*Oryza sativa*) (Wu *et al*., 2003) and a mutant population in wheat, produced by treatment of seeds with ethyl methanesulfonate (EMS) (Krasileva *et al*., 2017). There are limitations to all these techniques in that they are essentially random and will generate unwanted mutations, with chemical mutagenesis in particular producing a large number of 'background mutations'. This means that the sheer scale of the number of plants involved means that the method can be expensive and time-consuming, although it has been streamlined considerably in recent years with the availability of nucleotide sequence data of entire populations (Krasileva *et al*., 2017).

Methods have also been developed to target specific genes by genetic modification, including virus-induced gene silencing (Liu *et al*., 2002; Ruiz *et al*., 1998), and antisense RNA (Mol *et al*., 1990) and RNA interference (Smith *et al*., 2000). These methods, variously termed post-transcriptional gene silencing (PTGS), rely on interrupting the function of chosen genes by targeting their corresponding mRNA and, whilst they are used extensively in gene functional studies (Waterhouse and Helliwell, 2003), they also have their limitations, particularly if total loss of function is desired, as the suppression happens indirectly. As a result, incomplete or partial repression of a gene is often seen, resulting in a 'knockdown' in gene function rather than a 'knockout'. However, this can be beneficial if knockouts are lethal or deleterious.

1.12. Sequence-specific nucleases (SSNs)

The development of programmable, sequence-specific nucleases (SSNs) brought about the advent of a new era in gene functional studies and mutant generation. SSNs induce double-stranded breaks (DSBs) in specific targeted chromosomal sites, and rely on errors in the DNA repair pathways to generate mutations. There are two DNA repair pathways: the non-homologous end joining (NHEJ) pathway, which is the most common repair pathway, and the homology-directed repair (HDR) pathway (Lieber, 2010; Symington and Gautier, 2011; Puchta, 2005). The NHEJ pathway is particularly error prone, and often results in random insertions, deletions, and substitutions. This is because the NHEJ pathway relies on DNA ligase IV to repair the separated ends and, as there is no template involved, errors can occur if bases have been deleted by the nuclease (Belhaj *et al*., 2015). If

these occur in the coding region, these can result in frameshift mutations and gene knockouts. The HDR pathway repairs a DSB based on a homologous donor template, if one is present. This pathway can be used to achieve gene insertions or precise gene modifications. T-DNA from Agrobacterium (*Agrobacterium tumefaciens*)-mediated transformation has been shown to be preferentially integrated at DSBs in tobacco (*Nicotiana tabacum*) (Salomon and Puchta, 1998); however, few studies have shown the successful use of HDR-mediated genome editing in plants (Shukla *et al*., 2009; Schiml *et al*., 2014; Wang *et al*., 2017; Ali *et al*., 2020; Hahn *et al*., 2018; Vu *et al*., 2020).

The first generation of SSNs took the form of zinc finger nucleases (ZFNs) (Kim *et al*., 1996; Bibikova *et al*., 2002), and these have been used widely in plants (Zhang *et al*., 2010; Lloyd *et al*., 2005; Shukla *et al*., 2009; Osakabe *et al*., 2010; Cai *et al*., 2009; Wright *et al*., 2005). However, the constructs for ZFN are costly, difficult and time-consuming to generate (Ramirez *et al*., 2008). Subsequently, a second SSN system was adapted from *Xanthomonas* bacteria for plants, based on transcription activator-like effector nucleases (TALENs) (Moscou and Bogdanove, 2009; Boch *et al*., 2009; Christian *et al*., 2010; Li *et al*., 2011). Targeting of TALENs, although easier to manipulate than that of ZFNs, requires the construction of complex tandem repeat domains in the TAL proteins (Sanjana *et al*., 2012).

The most recent SSN system to be developed was the clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein9 (Cas9)-based genome editing tool: the CRISPR/Cas9 system. This system was adapted from the type II CRISPR adaptive immunity system in the bacterium *Streptococcus pyogenes* (Jinek *et al*., 2012), which protects bacteria and archaea from invading nucleic acids (such as those of viruses) by recognising and cleaving foreign DNA. The CRISPR/Cas9 genome editing system can be engineered to target specific regions in a genome through the development of a specific, single-guide RNA (gRNA) and Watson-Crick base-pairing between this gRNA and DNA. This makes it much easier to manipulate than the protein-based systems that require specific DNA-binding domains. The introduction of stable and heritable CRISPR/Cas9-induced mutations in plants has been confirmed by numerous reports (Jiang *et al*., 2013b; Feng *et al*., 2014; Zhang *et al*., 2014a; Zhou *et al*., 2014; Ma *et al*., 2015b; Li *et al*., 2013; Nekrasov *et al*., 2013; Shan *et al*., 2013; Xie *et al*., 2015).

The main advantages of CRISPR/Cas9 are its simplicity, accessibility, cost and versatility. ZFNs and TALENs are based on protein components and function as dimers, with specificity determined by their DNA-binding domains. ZFN design is complex and difficult due to the complicated nature of the interaction between zinc fingers and DNA (Sander *et al*., 2011). Editing efficiencies are generally higher in commercially available ZFNs, but these are more expensive (Ramirez *et al*., 2008). TALENs are easier to manipulate due to the one-to-one recognition rules between the TALENS protein repeats and nucleotide sequences, and DNA assembly techniques such as Golden Gate cloning

(Engler *et al*., 2008) have also helped improve construction. TALEN construction still has problems, however, with the highly repetitive sequences causing homologous recombination in human cells (Holkers *et al*., 2013). The CRISPR/Cas9 system consists of a single, monomeric protein and a chimeric RNA, with specificity determined by the gRNA target sequence, a 20-nucleotide sequence within the gRNA. In contrast to the other SSN systems, the CRISPR/Cas9 system is easier to manipulate as no sophisticated protein engineering is required. It is easy to change the target specificity of the system as the specificity is encoded by the simple 20 nucleotide gRNA target sequence.

In theory, ZFNs can target any sequence but, in reality, targeting is limited by the context-dependent interaction between the fingers, and high failure rates have been reported in construction (Sander *et al*., 2011). TALEN targets only require a thymine nucleotide at the first position (Doyle *et al*., 2012); however, not all TALENs work well, and some pairs do not generate mutations *in vivo*, requiring experimental validation of each TALEN pair that is designed. The CRISPR/Cas9 system requires the presence of a protospacer adjacent motif (PAM), which does limit the available target sites. However, the CRISPR system is easy to multiplex, allowing for the simultaneous editing of multiple targets (see Section 1.15). It can also introduce DSBs in methylated DNA in human cells (Hsu *et al*., 2013). When comparing the editing efficiencies of the SSN systems, the CRISPR/Cas9 system shows comparable, or even higher editing efficiencies than ZFNs or TALENs (Belhaj *et al*., 2015). In maize, for example, simultaneous editing of a target site by CRISPR/Cas9 and TALENs showed editing efficiencies of 13.1 % and 9.1 %, respectively (Liang *et al*., 2014). However, there are a lot of factors that affect CRISPR/Cas9 editing, so mutation rates can vary, even in the same species. In rice protoplasts, for example, editing of the *PDS* gene showed an editing efficiency of up to 38 % (Shan *et al*., 2013), whilst efforts to target the *MPK5* gene only showed an editing efficiency of 8 % (Xie and Yang, 2013).

1.13. Components of the CRISPR/Cas9 system

The CRISPR/Cas9 system has three components: the gRNA target sequence, the gRNA scaffold and the Cas9 nuclease (Figure 1.3). The gRNA requires a PAM sequence (5'-NGG-3' for the Cas9 from *Streptococcus pyogenes*) (Jinek *et al*., 2012; Sternberg *et al*., 2014) to be directly downstream of the target sequence. The sequence 5'-NAG-3' has been shown to function as a PAM but targeting efficiency was significantly reduced (Hsu *et al*., 2013).

The gRNA scaffold consists of the protospacer-containing CRISPR RNA (crRNA) and the *trans*-activating crRNA (tracrRNA), which greatly resembles the original bacterial system but has been simplified, with the gRNA target sequence situated at the 5′ terminal part (Jinek *et al*., 2012; Cong *et al*., 2013). In the bacterial immune system, the CRISPR arrays present in the genome are transcribed

and processed into CRISPR RNAs (crRNAs), which contain variable targeting sequences. These then activate and guide the Cas9 nuclease through interactions with the transactivating CRISPR RNA (tracrRNA) (Barrangou *et al*., 2007). The targeting specificity of the system is provided by the 20 bp region upstream of the PAM sequence, which complements the target sequence of the genomic DNA. The Cas9 nuclease shows a conserved, bi-lobed architecture (Jinek *et al*., 2014; Nishimasu *et al*., 2014) consisting of the small nuclease (NUC) lobe, containing adjacent active sites (the HNH and RuvC domains), as well as a PAM-interacting domain, and a larger recognition (REC) lobe, which interacts with the gRNA, forming the binary Cas9-sgRNA complex (Nishimasu *et al*., 2014)

The gRNA (scaffold + targeting sequence) forms a complex with the Cas9 nuclease, called the Cas9 nuclease complex (Jinek *et al*., 2012; Jiang *et al*., 2015; Nishimasu *et al*., 2014). The gRNA binds to the genomic target complimentary strand and the RucV and HNH domains of the Cas9 endonuclease cut the DNA, introducing a DSB with blunt ends. Multiple gRNAs can be designed with different target sequences to target multiple loci in the same cell (Cong *et al*., 2013), allowing for targeting of multiple genes, as well as whole gene families (Xie *et al*., 2015; Ma *et al*., 2015b; Sánchez-León *et al*., 2018).

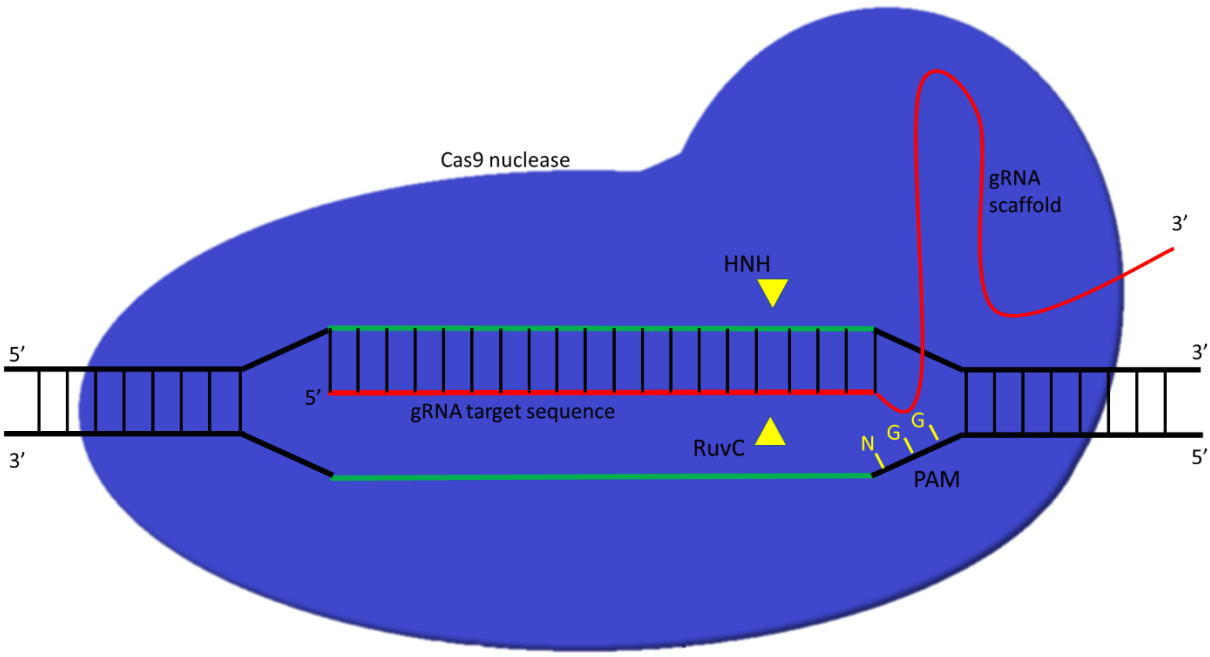


Figure 1.3. Illustration of the Cas9-gRNA complex. The Cas9 nuclease (blue) interacts with the DNA through the gRNA (red) and introduces double stranded breaks at the HNH and RucV cleavage sites (yellow triangles). The gRNA consists of a 20 bp target sequence (shown bound to the DNA) and the gRNA scaffold, which interacts with the Cas9 protein. The protospacer adjacent motif (PAM; 5'-NGG-3') is required for successful targeting.

1.14. CRISPR/Cas9 vectors in Plants

The CRISPR/Cas9 system was developed for human cells, and when tested in plants, although successful in principle (Li *et al*., 2013; Nekrasov *et al*., 2013; Shan *et al*., 2013; Xie and Yang, 2013), achieved much lower editing efficiencies than had been seen in other systems (Xing *et al*., 2014; Zhou *et al*., 2014). This led to the development of CRISPR/Cas9 vector systems specific to plant systems.

A gRNA typically contains 98 nucleotides, of which 20 correspond to the target sequence and guide the system to the genomic region of choice (Nishimasu *et al*., 2014). The original *Cas9* coding sequence is 4107 nucleotides long (Liu *et al*., 2015c), although a fused nuclear localization signal (NLS) is required in eukaryokes (Cong *et al*., 2013). The use of Cas9 in plants can show a low editing efficiency, as the gene comes from a prokaryotic system and was first modified to work in human systems; thus, plant-optimised *Cas9* genes have been designed, where the gene has been adjusted for plant codon bias (Jiang *et al*., 2013b; Li *et al*., 2013; Miao *et al*., 2013; Shan *et al*., 2013; Fauser *et al*., 2014; Xing *et al*., 2014; Svitashev *et al*., 2015; Gao *et al*., 2015; Ma *et al*., 2015b). In rice, the *Cas9* gene was further modified to mimic Gramineae genes (Wong *et al*., 2002) by increasing the GC content of the 5′ terminal region (Ma *et al*., 2015b). The *Cas9* gene is generally placed under the control of a constitutive promoter, such as the Ubiquitin (*Ubi*) promoter of maize, rice, or Arabidopsis, or the *Cauliflower mosaic virus* (*CaMV*) 35S promoter (reviewed in Ma *et al*., 2016).

The expression cassettes containing the *Cas9* and gRNA genes are delivered into the plant cells as plasmid constructs. The *Cas9* and gRNA genes can be placed on a single construct or delivered as separate constructs (Jiang *et al*., 2013b; Li *et al*., 2013; Nekrasov *et al*., 2013; Shan *et al*., 2013; Xie and Yang, 2013), and the plasmids carrying them are often tested first *in vitro* or in a transient expression system, such as protoplasts, or vacuum infiltrated plant leaves (Jiang *et al*., 2013b). It is not always possible to recover edited transgenic plants from protoplasts, and some species do not regenerate from protoplasts at all, including wheat (Vasil and Vasil, 1980); however, in vegetatively-propagated plants this method can be used effectively, with, for example, DNA-free genome edited plants generated from the transfection of Arabidopsis, a tobacco species (*Nicotiana attenuata*) and lettuce (*Lactuca sativa*) using preassembled CRISPR/Cas9 ribonucleoproteins (Woo *et al*., 2015). Plants have also been regenerated from rice protoplasts using this method (Woo *et al*., 2015). Most commonly, familiar plant transformation methods are used, such as the biolistic transformation of calli and immature embryos or Agrobacterium-mediated transformation (reviewed in Chapter 5), either through the floral dip method or through the transfection of callus, immature embryos, or other tissues.

1.15. Multiplexing the CRISPR/Cas9 system

CRISPR/Cas9 can be used for the simultaneous targeting of multiple genes, such as gene families, or multiple loci within one gene. This can require the use of multiple vectors, which becomes more difficult when using Agrobacterium-mediated transformation (Hiei and Komari, 2008). Furthermore, the use and construction of multiple expression cassettes can become tricky and problematic. These problems can be overcome with the use of multiplexed systems, whereby many loci can be targeted at once, or one locus targeted multiple times.

Multiple gRNA expression cassettes can be inserted into a single vector through sequential rounds of regular cloning (Li *et al*., 2013; Zhou *et al*., 2014). Another cloning method relies on the use of multiple restriction enzymes with compatible, palindromic 'sticky' ends to insert multiple gRNA expression cassettes into a single vector (Zhang *et al*., 2015).

The Golden Gate cloning method relies on distinctive type II restriction enzymes, such as *Bsa*I, to generate sequential, compatible, non-palindromic sticky ends, allowing for the simultaneous and efficient joining of multiple DNA fragments (Engler *et al*., 2008). Several CRISPR/Cas9 vector systems have been developed based on this process (Xing *et al*., 2014; Ma *et al*., 2015b). One strategy for gRNA cassette construction is the creation of multiple, independent, intermediate vectors which are then digested, and the excised gRNA cassettes are then ligated together *via* a Golden Gate reaction (Lowder *et al*., 2015). CRISPR/Cas9 constructs with up to eight gRNA expression cassettes have been produced (Ma *et al*., 2015b), leading to the simultaneous editing of seven *FT*-like genes in rice.

Another method is the Gibson assembly method, which joins multiple DNA fragments with homologous termini by using the T5 exonuclease, Phusion DNA polymerase, and *Taq* DNA ligase in concert (Gibson *et al*., 2009). This can generate vectors containing multiple gRNA expression cassettes (Ma *et al*., 2015b).

Multiple gRNAs can also be multiplexed into a single expression cassette through the polycistronic tRNA-gRNA (PTG) system (Xie *et al*., 2015). This system works by interspacing gRNAs with tRNA precursor sequences in a single gene, with the *in vivo* processing of the pre-tRNAs releasing the gRNAs from the polycistronic pre-RNA, allowing multiplexed targeting. This is discussed further in Chapter 4 (Section 4.1.3).

1.16. Factors affecting the efficiency of the CRISPR/Cas9 system

The editing efficiencies achieved in studies employing the CRISPR/Cas9 system in plants vary widely (Belhaj *et al*., 2015). Varying *Cas9* expression levels may explain this, with the codon optimisation of the *Cas9* gene (Wang *et al*., 2015), the promotor used (Yan *et al*., 2015), and the position where *Cas9* integrates into the host genome (Kohli *et al*., 2003) all factors that could affect this. Other *Cas9* genes

have been identified, such as *StCas9* (Xu *et al*., 2015) from *Streptococcus thermophilus*, and *SaCas9* (Ran *et al*., 2015), from *Staphylococcus aureus*, which have been shown to function in plants (Steinert *et al*., 2015; Jia *et al*., 2017). Another gene, *Cpf*, also adapted from the Class II CRISPR system, in which the effectors are single large proteins, has been shown to create DSBs in DNA, but with sticky ends (Zetsche *et al*., 2015).

Transient expression studies have implicated the gRNAs in regulating editing efficiency (Jinek *et al*., 2013; Li *et al*., 2013). This does not appear to be a product of gRNA expression levels, as different expression levels produced similar editing efficiencies in rice (Ma *et al*., 2015b). gRNA design may be a factor, and it used to be believed that the gRNA should start with a particular nucleotide (A for U3 promoters and G for U6 promoters (Shan *et al*., 2013)); however, it has since been shown that gRNAs with extra nucleotides, and guide sequences that do not begin with A or G, also lead to successful editing (Ran *et al*., 2013; Xie and Yang, 2013; Ma *et al*., 2015b), with editing efficiencies being similar to the regular gRNAs (Ma *et al*., 2015b). Target sequences with higher GC contents (50 – 70 %) have shown higher editing efficiencies in rice (Ma *et al*., 2015b), indicating that the target sequence composition may play a role in CRISPR/Cas9-induced editing efficiency (Zhang *et al*., 2014a; Ma *et al*., 2015b; Wang *et al*., 2013). The 3D structure may also affect the efficiency, as the formation of a stem-loop structure involving the target sequence and the scaffold affects binding to the genomic target strand (Ma *et al*., 2015b). In addition to this, the modification of the gRNA scaffold can improve editing efficiency, with Dang *et al*. (2015) showing that extending the duplex length, and mutating the polythymine sequence, increased editing efficiencies in human cell lines.

The efficiency of the system may also be affected by the genomic context of the target sequence, with epigenetic factors, such as DNA methylation or histone modification, affecting the ability of DNA-binding proteins to interact with the DNA (Tate and Bird, 1993). Chromatin immunoprecipitation and high-throughput sequencing (ChIP-seq) analysis has shown that Cas9 preferentially binds to open chromatin (Kuscu *et al*., 2014; Wu *et al*., 2014b); however, Cas9 has been shown to cleave methylated DNA (Hsu *et al*., 2013). The PAM sequence chosen also affects the editing efficiency. The PAM sequence 5'-CGG-3', for example, was shown to improve Cas9 specificity (Doench *et al*., 2014), especially when a base other than a guanine was adjacent to it (5'-CGGH-3').

Finally, cell type and the delivery method of the system appear to affect CRISPR/Cas9 activity. When targeting the *PDS* gene in tobacco, *Nicotiana benthamiana*, a mutation rate of 37.7 % was seen in PEG-mediated transfected mesophyll protoplasts, but only 4.8 % was seen when the same constructs were transfected *via* agroinfiltration into whole leaves (Li *et al*., 2013). This could be the result of differences in transformation efficiency, different DNA repair mechanisms in different cell types, or different gRNA/Cas9 expression levels.

1.17. Off-target effects

Eukaryotic genomes can be large, and often contain homologous and repeated sequences. Thus, off-target effects pose a genuine concern for the CRISPR/Cas9 system, with a high level of off-target effects reported in early studies (Fu *et al*., 2013; Hsu *et al*., 2013; Pattanayak *et al*., 2013).

Many web-based tools now exist for gRNA target selection (reviewed in Chapter 4), with algorithms that factor in off-targeting potential. Optimising the *Cas9* gene may also lead to increased targeting specificity by substituting residues involved in non-specific DNA contact (Kleinstiver *et al*., 2016; Slaymaker *et al*., 2016), using pairs of *Cas9*-derived nick-nuclease/sgRNAs (Ran *et al*., 2013) and changing the gRNA scaffold (Dang *et al*., 2015). This is discussed further in Chapter 4 (Section 4.1.5).

1.18. Types of mutations seen in the CRISPR/Cas9 system

Large mutations appear to be rare in the CRISPR/Cas9 system, and larger deletions more common than large insertions, with half the edits reported being single-base (mostly A and T) insertions and the rest small deletions (1–50 bp) (Feng *et al*., 2014; Zhang *et al*., 2014a; Ma *et al*., 2015b). Thus, one of the most common outcomes from genome editing is an event that causes a frameshift and results in loss-of-function of the gene. The use of multiple gRNAs to target a single locus can result in large deletions (up to hundreds of kilobases in length) between the gRNA target sites (Li *et al*., 2013; Zhou *et al*., 2014; Zhao *et al*., 2016). The CRISPR/Cas9 system can introduce biallelic or homozygous mutations in the first generation in rice and tomato (Brooks *et al*., 2014; Shan *et al*., 2013; Zhang *et al*., 2014a; Zhou *et al*., 2014). These edits are transferred to the following generations, and segregate normally, without further modifications. In plants with homozygous or biallelic mutations in the T0, further mutations were not seen in the subsequent generation, suggesting that once a target site has been edited, further editing does not occur (Feng *et al*., 2014).

1.19. Applications of CRISPR/Cas9 in plants

CRISPR/Cas9 provides a powerful tool for gene functional studies; however, it also has the potential to be a powerful tool in plant breeding and crop improvement. In particular, traits that have a negative effect on crop desirability, such as asparagine accumulation in wheat grain, could be improved by knocking out the genes involved, such as *TaASN2*. The simultaneous editing of the three *TaMLO* homologues in bread wheat, for example, improved resistance to powdery mildew (Wang *et al*., 2014), and the knockout of the ERF transcription factor gene *OsERF922* in rice reduced susceptibility to rice blast (Wang *et al*., 2016). Successful editing has been shown in a variety of

species, and a protocol for applying CRISPR/Cas9 in rice and wheat has been published (Shan *et al*., 2014).

Genome editing can be particularly useful in plants with long reproductive cycles (Bewg and Tsai, 2018); for example, the efficient editing of the phytoene desaturase gene 8 (*PtoPDS8)* gene in the Chinese white poplar (*Populus tomentosa Carr*.) (Fan *et al*., 2015). This gene has been targeted in many tree species; including citrus (*Citrus sinensis* cv. Valencia, Jia and Wang, 2014; *Poncirus trifoliata L. Raf.* × *Citrus sinensis L. Osb.*, Zhang *et al*., 2017; *Citrus paradisi* Macfadyen, Jia *et al*., 2017); coffee (*Coffea canephora*, Breitler *et al*., 2018); grape (*Vitis vinifera L.*, cv. Neo Muscat, Nakajima *et al*., 2017); and apple (*Malus prunifolia* (wild.) Borkh. 'Seishi' × *M. pumila* Mill. var. *paradisiaca* Schneid. 'M.9', Nishitani *et al*., 2016).

Targeted gene replacement, or gene knock-ins, through the HDR DNA repair pathway, is also a promising tool for crop improvement; however, the editing efficiency for HDR-mediated knock-ins is much lower than gene knockouts (Puchta and Fauser, 2014; Svitashev *et al*., 2015). Successful knock-ins using CRISPR/Cas9 have been reported in plants (Shukla *et al*., 2009; Schiml *et al*., 2014; Wang *et al*., 2017; Ali *et al*., 2020; Hahn *et al*., 2018; Vu *et al*., 2020; Li *et al*., 2015). The CRISPR/Cas9 system can also be applied for uses other than genome editing through the introduction of DSBs. Deactivated nucleases, such as dead Cas9 (dCas9), a catalytically inactive version of Cas9, can regulate gene expression as they will still bind to their target sequences. The dCas9 can be fused with a transactivation or transrepression domain to allow for precise and reversible transcriptional control of target genes (Gilbert *et al*., 2013; Maeder *et al*., 2013). The level of transcriptional control can be altered as using multiple gRNAs shows synergistic effects (Qi *et al*., 2013).

1.20. Aims of the project

Acrylamide is a Group 2a human carcinogen and a food processing contaminant. This thesis describes work to utilise genome editing technology, as well as investigating an EMS-mutagenised TILLING population, for the targeted knockout of asparagine synthetase 2 (*TaASN2*) in wheat, with the goal of reducing the acrylamide-forming potential of wheat.

The project focussed on the following objectives:

1. Analysis of the *TaASN* genes in wheat, leading to the design of effective gRNAs for the targeted knock-out of *TaASN2* through CRISPR/Cas9 technology.
2. The utilisation of TILLING mutant lines to generate *TaASN2* mutants.
3. Generation and characterisation of genome-edited plants with mutations in *TaASN2*.
4. Analysis of the *TaASN2*- knockout phenotype for plant growth and grain composition, and for asparagine accumulation.

Chapter 2

Materials and Methods

2.1. General materials and methods

2.1.1. Polymerase chain reaction (PCR)

Primers were designed using Primer3Plus (Untergasser *et al*., 2007) and used at a concentration of 10 pmol/µL. Two different polymerases were used: the 2 × ReddyMix PCR Master Mix (Thermo Fisher Scientific, Hemel Hempstead, UK) and the Q5 High-Fidelity DNA polymerase (New England Biolabs, Hitchin, UK), according to their recommended protocols. All PCR reactions were run as 25 µL or 50 µL reactions. To optimise the reactions, different annealing temperatures were trialled on a gradient setting.

Table 2.1. Primers used in PCR amplification of the *TaASN2* target site.

Primer name	Primer sequence
ASN2-1F	GATCCAGAGGAGCAGCATAACC
ASN2-2F	GTAGAGCCAAGCCATTCCTG
ASN2-1R	GCGATGACCTCGCAGTCAC

2.1.2. Gel electrophoresis

Gels contained TAE buffer (Tris-acetate- ethylenediaminetetraacetic acid (EDTA); 0.4 M tris acetate (pH approximately 8.3), 0.01 M EDTA) and agarose at a concentration of 1 %, 1.5 % or 2 %. Ethidium bromide was added as a DNA stain. Gels were visualised using a Bio-Rad GelDoc XR (Bio-Rad, Watford, UK). The GeneRuler 1 kbp, GeneRuler 100 bp (Thermo Fisher Scientific, Hemel Hempstead, UK) and 2-Log DNA (0.1 - 10.0 kbp) (New England Biolabs, Hitchin, UK) ladders were used as size markers.

2.1.3. DNA purification

All DNA purification from agarose gels and PCR reactions was performed using the 'Wizard® SV Gel and PCR Clean-Up System' (Promega, Southampton, UK), following the manufacturer's protocol. For DNA purification from agarose gels, the desired DNA fragments were cut from the gel and the gel slice was dissolved in an excess of the 'membrane binding solution' at 55 °C, before being added to the purification columns. For DNA purification from PCR reactions, an equal volume of the 'membrane binding solution' was added to the PCR products and then added directly to the purification columns. An additional 5-minute incubation at 60 °C following the wash cycles was added to the Level 1 PCR fragment gel extraction protocol. DNA quantification was performed using a NanoDrop 2000 (Thermo Fisher Scientific, Hemel Hempstead, UK).

2.1.4. Glasshouse conditions

Plants were grown in a containment glass house. They were watered daily and grown under a 16-hour day (with supplementary lighting when necessary), at a day temperature of 20 ºC and a night temperature of 15 ºC. Plants were self-pollinated and grain was collected at maturity.

2.1.5. Leaf sampling method

Leaf sections of approximately 2 cm were cut from the tips of the leaves, making sure to take from the oldest leaves of the seedlings to avoid plant death. The leaf sections were placed into 1.5 mL microcentrifuge tubes and frozen immediately in liquid nitrogen. The scissors were cleaned between samples to ensure that there was no cross-contamination. The samples were stored in a -80 °C freezer. The leaf samples were ground to a fine powder under liquid nitrogen. The liquid nitrogen was allowed to evaporate, and the samples were transferred back to the microcentrifuge tube they were being stored in. Fresh grinders were used every time.

2.1.6. DNA isolation from leaf samples

This method is based on the Wizard® Genomic DNA Purification Kit. Nuclei lysis solution (600 µL) (Promega A7943) was added to each ground leaf sample and the microcentrifuge tubes were vortexed for a few seconds. The samples were incubated at 65 °C for 15 minutes. RNAse A solution (1.5 µL, 10 mg/mL) was added, the tube was mixed by inversion and incubated at 37 °C for 15 minutes. The samples were then cooled to room temperature. Protein precipitation solution (200 µL) (Promega A7953) was added and the tubes were vortexed at high speed for 20 seconds. The samples were then centrifuged at 13,000 - 16,000 × g for 3 minutes, allowing the precipitated proteins to form a tight pellet. The supernatant was transferred, taking care not to disturb the protein pellet, to a fresh 1.5 mL microcentrifuge tube containing 600 µL of room temperature isopropanol. The DNA was washed by carefully inverting the tube and then the tube was centrifuged at 13,000 - 16,000 × g for 1 minute. The supernatant was decanted and 600 µL of room temperature 70 % ethanol added. The tube was inverted and centrifuged again. The ethanol was carefully aspirated to ensure the pellet was not disturbed. The excess ethanol was removed by inverting the tube onto absorbent paper, allowing the pellet to air dry for 15 minutes. H_2O (50 µL) was added to resuspend the pellet and the samples were incubated at 65 °C for 1 hour or left overnight at room temperature. The DNA was then stored at -20 °C.

2.2. Chapter 3 materials and methods

2.2.1. Investigating the *TaASN* genes in cv. Chinese Spring and cv. Cadenza

DeCypher Tera-BLASTN Search Nucleic Query vs. Nucleic Database was used to assess the wheat asparagine synthetase gene sequences. The NR_Gene_v0.4 scaffold wheat genome (PLANT_T.aestivum_NRgene_v0.4_scaf) was used first, due to its availability, and the cDNA nucleotide sequences for *TaASN1* (GenBank BT009245), *TaASN2* (GenBank BT009049), and *TaASN3* (GenBank AK333183) were used as the query sequences (Gao *et al*., 2016).

The returned scaffolds were downloaded and aligned to the cDNAs using the Geneious 10.1.3 software package (Kearse *et al*., 2012). Pairwise alignment was run using the Geneious Alignment algorithm on its default settings; multiple alignments were run using the Consensus Align algorithm, again on its default settings. The returned scaffolds were aligned to the cDNAs to identify exons and introns, and the results were later confirmed by searches of the *Triticum aestivum* IWGSC RefSeq v1.0 genomic sequence (Alaux *et al*., 2018), as it became available. The *Triticum aestivum* TGAC v1.0 database was also used to assess chromosomal positioning. *TaASN4* was identified through its divergence from the other wheat asparagine synthetase sequences as *TaASN4* has not yet been cloned from bread wheat.

The *ASN* family was investigated in the Cadenza genome using the DeCypher Tera-BLASTN Search Nucleic Query vs. Nucleic Database using Cadenza_EI_v.1 as the reference database. The database was first searched using the genomic DNA sequence of *ASN2* (from chromosome 3B), and subsequently with the other Chinese Spring *ASN* genomic sequences. The returned scaffolds were downloaded and aligned using the Geneious 10.1.3 software package (as above).

2.2.2. RNA-Seq analysis

The methods of this have been published in:

> Curtis, T.Y., Raffan, S., Wan, Y., King, R., Gonzalez-Uriarte, A. and Halford, N.G., 2019. Contrasting gene expression patterns in grain of high and low asparagine wheat genotypes in response to sulphur supply. *BMC Genomics*, *20*(1), p.628.

I analysed the dataset, identified genes of interest and produced the graphs and heatmap used in the figures.

2.2.3. Phylogenetic analysis of cereal asparagine synthetase genes

The EnsemblPlants database (Aken *et al*., 2017) was used to provide genome information for the species. The genome data for each chosen plant species was searched for genes annotated as expressing a protein containing an asparagine synthetase domain. Genes that had not been annotated previously were identified by BLASTn or BLASTp searches (Altschul *et al*., 1990) performed using the cDNA or derived amino acid sequence from wheat asparagine synthetase genes. The cDNA sequences were then downloaded and viewed using Geneious (Geneious 10.1.3) (Kearse *et al*., 2012).

The nucleotide sequences were trimmed so that regions upstream of the translation start codon and downstream of the translation stop codon were discarded. The nucleotide sequences were then aligned using Geneious Alignment on its default settings, with a cost matrix of 65 % similarity using a global alignment with free end gaps. The similarities between the genes were ascertained by the nucleotide alignment matrix. All cDNA alignments were confirmed using MUSCLE alignments within the Geneious package (Kearse *et al*., 2012). The similarity values between the asparagine synthetase genes for a single species were calculated from individual alignments of those genes, with a second independent alignment generating the similarity values in comparison to the wheat genes.

The alignments were used to build trees, using Geneious tree builder software (Kearse *et al*., 2012), to visualise the relationship between the genes. The Jukes-Cantor model was used for genetic distance, and the tree was built *via* a Neighbor-Joining method. The Jukes-Cantor model is a simple substitution model which assumes that all bases have the same equilibrium base frequency and that nucleotide substitutions occur at equal rates (Jukes and Cantor, 1969). The Neighbor-Joining method was employed as it allowed for unrooted trees to be generated quickly without assuming a molecular clock (Saitou and Nei, 1987). A re-sampling method was also employed. Bootstrapping (Felsenstein, 1985) was performed with 100 replicates, creating consensus trees, which were then used to examine the relationship of the genes across the cereal species. The consensus tree allowed for an estimation of the support for each clade, and a 50 % threshold was applied in a majority rule consensus tree. The length of the branches in the consensus tree corresponded to the average over all trees containing the clade, with the length of the tip branches calculated by averaging over all trees. For the tree comparing all the *ASNs* in this study, an Arabidopsis *ASN* gene, *AtASN1* (Lam *et al*., 1994), was used as an outgroup to anchor the tree. The scale bars on the tree represented the length of the branches and were expressed in units of substitutions per site of the sequence alignment.

2.3. Chapter 4 materials and methods

2.3.1. gRNA generation

Potential gRNAs were initially designed using a *TaASN2* cv. Chinese Spring consensus sequence generated from the NRGene v4 database (Wilkinson *et al*., 2016) using the Sequencher 5.2.4 software package (Gene Codes Corporation, Ann Arbour, Michigan, US) using its basic alignment algorithm. Initially, a gRNA-generator (CerealsDB, University of Bristol, unpublished data) was used and the gRNA viability was visually assessed by alignment to the homeologue sequences. Further gRNA design was completed using the Cas9 Designer tool (Park *et al*., 2015) using '*Hordeum vulgare* (Ensembl Plants 28)' as the reference sequence because a wheat reference genome was not available for this tool at the time. The DESKGEN (Hough *et al*., 2016) gRNA design tool was also used with

gRNAs designed based on the *TaASN2* chromosome 3A genomic sequence from Cadenza (Appels *et al*., 2018).

The potential gRNAs were aligned to the cv. Cadenza *TaASN2* genes and scored based upon their GC contents, out-of-frame scores (a prediction of how likely a guide is to lead to a frameshift), and whether they targeted a conserved region between the homeologues. The gRNAs were also aligned to the cv. Cadenza *TaASN1* sequences (on Chromosome 5A, 5B and 5D) to assess the likelihood of off-target effects in any conserved regions between *TaASN1* and *TaASN2*. The online tool DESKGEN (Hough *et al*., 2016) was used to confirm the viability of the gRNAs, through the 'knock-in function', with the genomic sequence of *TaASN2* from Chromosome 3A of Cadenza used as the region of choice. The 3D structure of the gRNAs was visualised through RNA-fold (Lorenz *et al*., 2011) as a further check of gRNA suitability.

2.3.2. Creation of Actin-Cas9-eGFP plasmid

The Actin-Cas9-eGFP plasmid was created through the two-step digestion of the destination plasmid, pRRes208.381 (obtained from Alison Huttly, Rothamsted Research), and the donor plasmid (containing a *Cas9* construct with the maize ubiquitin (*Ubi*) gene promoter, obtained from the University of Bristol) with *Eco*R1 and *Spe*1 restriction enzymes (New England Biolabs, Hitchin, UK), according to the manufacturer's protocol, and the ligation of the digested fragments. The DNA fragments were purified using the QIAquick PCR Purification Kit (Qiagen, Manchester, UK), following the manufacturer's protocol, and eluted in 30 µL elution buffer (10 mM Tris-Cl, pH 8.5). Recombinant shrimp alkaline phosphatase (3 µL) (New England Biolabs, Hitchin, UK) was added to the pRRes208.381 plasmid digestion one hour after the start of the *Spe*I digestion to catalyse the dephosphorylation of 5´ and 3´ ends of the DNA. The ligation reaction set up is shown in Table 2.2.

Table 2.2. Ligation reaction components for the creation of the Actin-Cas9-eGFP plasmid. Five different ligation mixes were set up, all as 20 µL reactions.

	Quantity (µL)				
Ligation Mix	Vector	Insert	10× buffer	Ligase	H_2O
1	2	-	2	-	16
2	2	-	2	2	14
3	2	5	2	-	11
4	2	5	2	2	9
5	2	10	2	2	4

The ligation reaction was incubated overnight at room temperature, then stored at -20 °C.

2.3.3. *Escherichia coli* (*E. coli*) cells

Chemically competent *E.coli* strain DH5α cells (produced at Rothamsted Research by Lucy Hyde) and 10-beta cells (New England Biolabs, Hitchin, UK) were used. The cells were stored at -80 °C and transformed *via* a 'heat shock' procedure as follows: The cells were removed from the -80 °C freezer and incubated on ice for 5 minutes. DNA was added and the cells were incubated on ice for a further 30 minutes. Cells were then transferred to a 42 °C water bath for 30 seconds and immediately returned to ice and incubated for 5 minutes. SOC medium (Sigma-Aldrich, Gillingham, UK) was added and the cells were transferred to a 37 °C shaking incubator and incubated at 150 - 200 rpm for 1 hour. Cells were removed from the incubator and were grown on YT medium (16 g/L tryptone, 10 g/L yeast extract; 5.0 g/L NaCl) containing a selective antibiotic and grown overnight at 37 °C.

2.3.4. Polymerase chain reaction (PCR) analysis of *E. coli* colonies

Colonies were streaked on selective plates containing ampicillin (100 µg/mL) to ensure that single colonies were obtained. These colonies were then added to 100 µL of EB buffer (10 mM Tris-Cl, pH 8.5) and incubated at 95 °C for 10 minutes. PCR reactions were run as 25 µL reactions, where the EB mix was used as the template DNA, using the following programme: 95 °C for 5 minutes, [95 °C for 30 seconds, 58 °C for 30 seconds, 72 °C for 60 seconds] × 25 cycles, 72 °C for 10 minutes. 'Positive' colonies were confirmed by nucleotide sequence analysis and used to inoculate LB broth (tryptone 10 g/L; NaCl 10 g/L; yeast extract 5 g/L) to generate plasmid DNA midipreps (using the Qiagen Plasmid Midiprep Kit (Qiagen, Manchester, UK), following the manufacturer's protocol).

2.3.5. Nucleotide sequence analysis

Nucleotide sequence analysis of protoplast DNA was conducted on an Applied Biosystems® 3500 Genetic Analyzer at the University of Bristol. Results were downloaded and analysed using either Sequencher or Geneious software (Gene Codes Corporation, Ann Arbour, Michigan, US; Kearse *et al*., 2012). The primers used in the sequencing reactions were: Actin-1F[1], Actin-1R[1], Histone-1F[1], Histone-1R[1], Nos2-R[1], ASN2-1F (GATCCAGAGGAGCAGCATAACC) and ASN2-1R (GCGATGACCTCGCAGTCAC) ([1]primers were obtained from Professor Keith Edwards, University of Bristol). Reactions were run in 10 µL volumes on a thermal cycler (Mastercycler nexus, Eppendorf, Stevenage, UK). The DNA was then precipitated with ethanol and resuspended in 10 µL of formamide loading buffer (95 % formamide). All additional nucleotide sequence analysis was performed by the Eurofins Custom DNA Sequencing service (Wolverhampton, UK), which uses Sanger sequencing.

2.3.6. Golden Gate assembly of plasmid constructs

The method was based on the protocol by Xie *et al*. (2015). Primers were designed to amplify the gRNA-tRNA components (Table 2.3). The system requires multiple DNA components to be ligated together in a desired order, and for Golden Gate (GG) assembly (Engler *et al*., 2008). Distinct 4-bp overhangs are required in order to ligate two DNA fragments after digestion with a type II endonuclease (*Bsa*I in this case). As the gRNA spacer is the only unique sequence, the full polycistronic gene should be designed in the form of DNA fragments where the overlaps are within the gRNA spacer region. The overhang could be any 4 consecutive nucleotides of the gRNA spacer; however, the spacer could not be 5'-GGCA-3' or 5'-AAAC-3', which are used in terminal adaptors for cloning in pRGE32 and pRGEB32, or 5'-GGATG-3' or 5'-GGTCTCN-3', which are the *Fok*I and *Bsa*I restriction sites, respectively. The gRNA spacers were split into two parts with a 4 bp overlap and each half of the spacer was synthesized within oligonucleotide primers with a *Bsa*I site. Two 5'-terminal bases were added to enhance *Bsa*I digestion of the PCR products. The gRNA scaffold was improved as suggested by Dang *et al*. (2015), changing the gRNA[x]-F primer to 5'-NN-GGTCTC-N-$N_9N_{10}N_{11}N_{12}N_{13}N_{14}N_{15}N_{16}N_{17}N_{18}N_{19}N_{20}$-GTTTCAGAGCTATGC-3', with the gRNA[x]-R primer remaining unchanged, 5'-NN-GGTCTC-N- $N_{12}N_{11}N_{10}$ $N_9N_8N_7N_6N_5N_4N_3N_2N_1$-TGCACCAGCCGGG-3' (Xie *et al*., 2015).

Table 2.3. Primers for the GG assembly of the 4-gRNA polycistronic gene, designed according to the method outlined in Xie *et al*. (2015).

Primer name	Primer sequence
L5AD5-F	CGGGTCTCAGGCAGGATGGGCAGTCTGGGCAACAAAGCACCAGTGG
gRNA1 R	TAGGTCTCATCGCCGCACCCCTGCACCAGCCGGGAA
gRNA1 F	CGGGTCTCCGCGACGAGTCGCGTTTCAGAGCTATGC
gRNA2 R	TAGGTCTCACCGCTCCAGTCCTGCACCAGCCGGGAA
gRNA2 F	CGGGTCTCCGCGGCCTGCACCGTTTCAGAGCTATGC
gRNA3 R	TAGGTCTCACAGCCGCTCTACTGCACCAGCCGGGAA
gRNA3 F	CGGGTCTCCGCTGGTCGCCGGGTTTCAGAGCTATGC
gRNA4 R	TAGGTCTCAGTGACTGCGAGGTGCACCAGCCGGGAA
gRNA4 F	CGGGTCTCCTCACTGCCGGTCGTTTCAGAGCTATGC
L3AD5-R	TAGGTCTCCAAACGGATGAGCGACAGCAAACAAAAAAAAAAACACCGACTCG
S5AD5-F	CG GGTCTCAGGCAGGATGGGCAGTCTGGGCA
S3AD5-R	TAGGTCTCCAAACGGATGAGCGACAGCAAAC

The Level 1 PCR was performed on the pGTR template, with improved gRNA structure (tRNA (shown in red below) and gRNA scaffold (shown in blue)):

TTGGAGACCGAGGTCTCGGTTTCAGAGCTATGCTGGAAACAGCATAGCAAGTTGAAATAAGGCTAGTCCGTTA
TCAACTTGAAAAAGTGGCACCGAGTCGGTGCAACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCTGCC
ACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCATGGCAGAA

2.3.7. Construction of PTG gene

PCR reactions (50 µL) were set up to amplify DNA gRNA fragments for the Level 1 GG assembly for PTG construction. Q5® High-Fidelity 2 × Master Mix (New England Biolabs, Hitchin, UK) was used following the manufacturer's protocol. Forward and reverse primers (Table 2.3) were used at a final concentration of 0.5 µM, with 0.15 ng of pGTR template (Xie *et al*., 2015). The 2 × ReddyMix PCR Master Mix (Thermo Fisher Scientific, Hemel Hempstead, UK) was also used.

The PCR was run on the following programme: 98 °C for 2 minutes; then 35 × [98 °C for 10 seconds; 50 °C for 20 seconds; 72 °C for 20 seconds]; then 72 °C for 2.5 minutes. The samples were then held at 10 °C.

The PCR products were purified with the Wizard® SV Gel and PCR Clean-Up System (Promega UK, Chilworth, UK). The individual fragments were ligated together in a GG reaction to generate the plasmid construct pRRes209.481.ASN2 and 25 ng of each Level 1 fragment was used in the GG reaction (total volume 20 µL). The GG reaction was set up as shown in Table 2.4.

Table 2.4. Components for the GG ligation of the Level 1 PCR fragments into the 4-gRNA polycistronic gene.

Component	Amount (µL)
Level 1 fragments (25 ng/µL)	1 of each
Bovine Serum Albumin (1 ng/µL)	2
*Bsa*1 (New England Biolabs, Hitchin, UK)	0.5
T7 DNA ligase (New England Biolabs, Hitchin, UK)	0.5
2 × T7 DNA ligase buffer (New England Biolabs, Hitchin, UK)	10
Nuclease-free H_2O	To 20 µL

GG reactions were performed in a thermal cycler (Bio-Rad) under the following programme: 50 cycles × [37 °C for 5 minutes; 25 °C for 10 minutes]; then 20 °C for 1 minute. The samples were then held at 8 °C.

The GG product was then reamplified by PCR in a 50 µL reaction. The ligation product was used at 1:10 dilution (1 µL) and 2 × ReddyMix PCR Master Mix (Thermo Fisher Scientific, Hemel Hempstead, UK) was used with the S5AD5-F and S3AD5-R (10 µM) primers (Table 2.3). The PCR was run on the following programme: 95 °C for 2 minutes; then 35 cycles × [95 °C for 10 seconds; 60

°C for 20 seconds; 72 °C for 1 minute/kbp]; then 72 °C for 5 minutes. The samples were then held at 4 °C.

The PCR products were purified with the Wizard® SV Gel and PCR Clean-Up System (Promega UK, Chilworth, UK) and digested with *Fok*1 (New England Biolabs, Hitchin, UK) in 30 µL reactions using the Cutsmart buffer. The reaction was left overnight at 37 °C.

The fragments generated by the *Fok*1 digest were separated using gel electrophoresis and the DNA fragments of the expected size were cut from the gel and purified with the Wizard® SV Gel and PCR Clean-Up System (Promega UK, Chilworth, UK).

The ligation reaction to insert the PCR product into the vector (pRGE32) was set up in 10 µL volumes using the T4 ligase and the 5 × T4 ligase buffer (New England Biolabs, Hitchin, UK). Two different ligation reactions were set up, with different vector to insert ratios. Reaction 1 contained a 5:3 ratio of vector to insert whilst Reaction 2 contained a 5:2 ratio.

The ligation reaction was incubated at room temperature for at least 4 hours, and then frozen at -20 °C. The ligation product, pRRes209.481.ASN2, was used to transform *E. coli* DH5α cells using the heat shock procedure and the construct was confirmed *via* a diagnostic restriction digest, using *Mlu*1 and *Pst*1 in 3.1 Buffer (New England Biolabs, Hitchin, UK).

2.3.8. Protoplast generation

The method was based on a modified Shan *et al*. (2014) protocol. Protoplasts were generated from 9 - 12 day old, dark-grown wheat cv. Cadenza seedlings.

Mannitol solution (50 mL, 1 M) was made fresh, vortexed and heated to 55 °C to dissolve. Fresh enzyme solution (20 mL) was prepared (1.5 % cellulase RS (Sigma C0615); 0.75 % macerozyme R10 (Calbiochem); 10 mM KCL; 0.6 M mannitol; 20 mM MES(2-(N-morpholino)ethanesulfonic acid)-KOH, pH 5.6; made up to 20 mL with H_2O) and heated in a water bath to 55 °C for 10 minutes to inactivate proteases and enhance enzyme solubility, then allowed to cool. Once cool, 200 µL of 1 M $CaCl_2$ and 2 mL of 10 mg/mL bovine serum albumin were added.

Shoots (approximately 10 - 20) were placed in a Petri dish and cut into 0.5 - 1 mm strips using a fresh razor blade. Plasmolysis buffer (0.6 M mannitol) (10 mL) was added to the cut shoots and the shoots were incubated in the dark for 10 minutes. The plasmolysis buffer was removed, and all 20 mL of the enzyme solution was added. The cells were then incubated in the dark in a desiccator for 30 minutes, after which they were incubated for 4 - 6 hours (in the dark) at 23 °C with shaking at 60 rpm.

W5 solution (125 mM $CaCl_2$, 154 mM NaCl, 2 mM MES-KOH, 5 mM KCL) (50 mL) was placed on ice, while keeping another 50 mL of W5 at room temperature. After the incubation, 20 mL of (room

temperature) W5 was added to the cells and gently mixed. The cells were then filtered through a 40 µm nylon mesh and washed with the remaining 30 mL of W5. The cells were centrifuged at 80 g for 3 minutes (Thermo Fisher Scientific Multifuge X3R centrifuge) and resuspended in 3 mL of (cold) W5 solution. The cells were incubated for 30 minutes on ice and the cell concentration was assessed using a Neubauer haemocytometer (Figure 2.1). The W5 solution was removed and the cells were resuspended in MMG (0.4 M mannitol, 15 mM $MgCl_2$, 4 mM MES-KOH) to give a concentration of 1 million cells per mL.

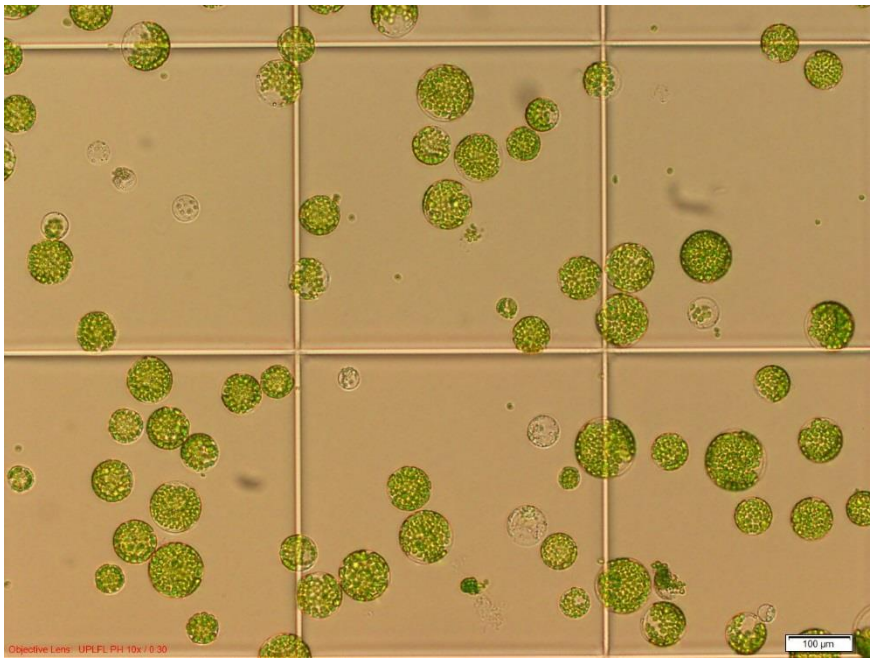


Figure 2.1. Protoplasts from cv. Cadenza viewed using a Neubauer haemocytometer to count for cell number.

For the transformation, 10 µg of plasmid DNA was used. Protoplasts (100 µL) were added to the plasmid DNA and mixed gently. PEG-CTS (110 µL) (40 % w/v polyethylene glycol (PEG), 0.2 M mannitol, 100 mM $CaCl_2$) was added and the cells were incubated at 23 °C for 10 minutes in the dark. The transformation was stopped by the addition of 880 µL W5 solution. The cells were centrifuged at 80 g for 3 minutes, resuspended in 2 mL of W5 and incubated in the dark for 24 - 48 hours. The effectiveness of the transformation was visualised through the addition of a fluorescent marker gene and fluorescence microscopy (M205 Leica microscope at 160× magnification). Some of the supernatant was removed (leaving 200 - 300 µL), concentrating the protoplasts prior to viewing. Transformation efficiency was estimated from the number of fluorescent to non-fluorescent living protoplasts.

2.3.9. gRNA1 and gRNA2 for protoplast trial assay

Two plasmids (pEXA24F-CONSENSUS.sgRNA; pEXA25F-CONSENSUS.sgRNA) were designed by Professor Keith Edwards (University of Bristol), each containing a single gRNA gene, using the Eurofins Gene Synthesis service, each comprising the pEX-A2 backbone and the following genes (with gRNA target sequence (in red below) and gRNA scaffold (in blue)):

gRNA1

GGATCCAAGCTTAAGAACGAACTAAACGCGTGCCGGACAAAAAAAGGAGCACATATACAAACCG
GTTTTATTCATGAATGGTCACGATGGATGATGGGGCTCAGACTTGAGCTACGAGGCCGCAGGCGA
GAGAAGCCTAGTGTGCTCTCTGCTTGTTTGGGCCGTAACGGAGGATACGGCCGACGAGCGTGTACT
ACCGCGCGGGATGCCGCTGGGCGCTGCGGGGGCCGTTGGATGGGGATCGGTGGGTCGCGGGAGCG
TTGAGGGGAGACAGGTTTAGTACCACCTCGCCTACCGAACAATGAAGAACCCACCTTATAACCCC
GCGCGCTGCCGCTTGTGTTGGGGTGCGGCGACGAGTCGCGTTTTAGAGCTAGAAATAGCAAGTTA
AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTTGAGATTTC
CAACCAGGTCCCTGGAGCCCATAGTCTAGTAACGGCCGCC

gRNA2

GGATCCAAGCTTAAGAACGAACTAAACGCGTGCCGGACAAAAAAAGGAGCACATATACAAACCG
GTTTTATTCATGAATGGTCACGATGGATGATGGGGCTCAGACTTGAGCTACGAGGCCGCAGGCGA
GAGAAGCCTAGTGTGCTCTCTGCTTGTTTGGGCCGTAACGGAGGATACGGCCGACGAGCGTGTACT
ACCGCGCGGGATGCCGCTGGGCGCTGCGGGGGCCGTTGGATGGGGATCGGTGGGTCGCGGGAGCG
TTGAGGGGAGACAGGTTTAGTACCACCTCGCCTACCGAACAATGAAGAACCCACCTTATAACCCC
GCGCGCTGCCGCTTGTGTTGGACTGGAGCGGCCTGCACCGTTTTAGAGCTAGAAATAGCAAGTTAA
AATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTTGAGATTTCC
AACCAGGTCCCTGGAGCCCATAGTCTAGTAACGGCCGCC

2.3.10. Protoplast DNA extraction Protocol

DNA was extracted from the protoplasts 48 hours after transformation. The extraction buffer (0.1 M Tris-HCl, pH 7; 0.05 M EDTA, 1.25 % sodium dodecyl sulphate (SDS); made up to 1 L with H_2O) was heated to 55 °C. The protoplasts were concentrated (i.e. most of the supernatant was removed, leaving 200 - 300 µL), 600 µL of extraction buffer was added and the cells were mixed vigorously and then incubated at 55 °C for 20 minutes. The cells were cooled to room temperature (approximately 20 °C) and 300 µL of 6 M ammonium acetate was added. The cells were left for 15 minutes at 4 °C, then centrifuged at top speed (14,500 rpm) in an Eppendorf™ MiniSpin™ plus Benchtop Centrifuge for 15 minutes before 600 µL of the supernatant was recovered. Isopropanol (360 µL) was added to the recovered supernatant in a fresh 1.5 mL microcentrifuge tube, mixed vigorously, and the DNA was left to precipitate for 5 minutes.

The samples were centrifuged again at 14,500 rpm for 10 minutes; the supernatant was discarded, and the pellet was washed with 400 µL of 70 % ethanol. The samples were centrifuged again at 14,500 rpm for 5 minutes, the ethanol was discarded, and the samples were then incubated at 65 °C for a further 5 minutes to ensure all residual ethanol had evaporated. The DNA was resuspended in 10 mM Tris-Cl (pH 8.5) or sterile H_2O, and either heated to 50 °C for 10 minutes and mixed *via* pipetting, or left overnight at 4 °C to allow the DNA to dissolve.

A PCR (25 µL reaction) using the primers ASN2-F2 and ASN2-1R (10 µM) and the 2 × Q5 buffer (New England Biolabs, Hitchin, UK) was set up to amplify the targeted region of *TaASN2*.

2.3.11. TOPO-cloning for the detection of edits in protoplast DNA

The PCR amplicons of the first exon of *TaASN2* extracted from the transformed protoplast population were inserted into the plasmid vector pCR™-Blunt II-TOPO™ using the Zero Blunt™ TOPO™ PCR Cloning Kit (Thermo Fisher Scientific, Hemel Hempstead, UK), as per the manufacturer's protocol, with 4 µL of amplicon, 1 µL of salt solution and 1 µL of PCR™-BLUNTII -TOPO™. Integration of the PCR amplicon disrupted the lethal *ccdB* gene present in the plasmid vector, allowing for selection of cells carrying recombinant plasmids.

The reaction was incubated at room temperature for 5 minutes and then placed on ice and used to transform One Shot™ *E. coli* cells following a heat-shock procedure (above). The cells were grown on YT medium containing kanamycin (100 µg/mL) and incubated overnight at 37 °C. Nucleotide sequence analysis of the plasmids was carried out using the M13 Forward and M13 Reverse primers provided with the kit.

2.4. Chapter 5 materials and methods

2.4.1. Plant Transformation

The transformation of immature cv. Cadenza embryos was performed by Lucy Hyde at the Cereal Transformation Group at Rothamsted Research (caroline.sparks@rothamsted.ac.uk). Three plasmids (pRRes209.481.ASN2, pRRes217.486 and pRRes111.1) were co-bombarded into the embryos through biolistic bombardment using a 'Gene Gun' (PDS-1000/He; Bio Rad, UK). The embryos were transferred to a selective medium containing phosphinothrycin after one week and then transferred to fresh medium every two weeks until the plants were established. The plants were then transferred to seed trays, with a cell size of height 5 cm × width 4.5 cm × length 4.5 cm and grown in a containment glass house.

2.4.2. Nucleotide sequence analysis

The nucleotide sequence analysis of the two T0 samples was performed by the Eurofins Custom DNA Sequencing service (Wolverhampton, UK), which uses Sanger sequencing.

2.4.3. TILLING population

The methods used with the TILLING population have been published on http://www.wheat-tilling.com/ under 'Additional resources and documentation of the website and populations'.

The http://www.wheat-tilling.com/ website was searched using the cv. Cadenza cDNA sequences for the *TaASN2* genes (TraesCAD_scaffold_023055_01G000100, TraesCAD_scaffold_036944_01G000100 and TraesCAD_scaffold_017129_01G000200 in the TGAC_cadenza, INSDC Assembly GCA_902810645.1, Dec 2015, accessed on Ensembl Plants). The output was downloaded and searched for mutations causing the formation of in-frame stop codons because these mutations are most likely to render the gene dysfunctional. The mutations were input into Geneious (software package 10.1.3 (Kearse *et al*., 2012)) and assessed. Mutations towards the 5' end of the gene were preferentially chosen, and homozygous lines were chosen over heterozygous lines.

2.5. Chapter 6 materials and methods

2.6.1. Next Generation Sequencing nucleotide sequence analysis

Next Generation Sequencing (MiSeq v2 2x250 bp Nano; Illumina TruSeq® Nano DNA) and analysis was performed at the University of Bristol by the Bristol Genomics Facility and Paul Wilkinson. The editing target region of the *TaASN2* gene was amplified from leaf DNA by PCR using primers 5'-GTAGAGCCAAGCCATTCCTG and 5'- GCGATGACCTCGCAGTCAC. Adapters comprising 4 bps were added at the 5' end to act as barcodes, with a unique barcode for each plant (Table 2.5). NGS was performed using a MiSeq v2 Benchtop Sequencing System (Illumina, Cambridge, UK) at the Bristol Genomics Facility (University of Bristol) using TruSeq DNA Nano Kits and MiSeq Reagent Kits v2 Nano (Illumina). The adaptors allowed for multiplexing in the NGS analyses because reads could be assigned to each plant using the barcodes.

PANDAseq version 2.11 (Masella *et al*., 2012) was used to assemble the original fastq paired reads into contigs and demultiplexed by plant using a bespoke perl script. The fasta file generated for each plant was aligned to the reference sequence using a Bowtie 2 build version 2.3.4.1 (Langmead and Salzberg, 2012). A read number of less than 200 was automatically discounted as, with six alleles/homeoalleles present, this would not give sufficient information on all of the *TaASN2* gene nucleotide sequences present.

Table 2.5. Primer adaptors used in the T1 and T2 NGS analyses. The 4-base adaptors were added in between the target primers (forward (GTAGAGCCAAGCCATTCCTG) and reverse (GCGATGACCTCGCAGTCAC)) and the Illumina TruSeq adaptors to allow for sample barcoding and for the DNA to be pooled in the analysis. Fewer adaptors were used in the T2 analysis due to fewer plants being analysed.

Forward Adaptors				Reverse Adaptors			
		Analysis				Analysis	
Name	Sequence	T1	T2	Name	Sequence	T1	T2
1F	AATT	X	X	1R	AACC	X	
2F	ATTT	X	X	2R	ACCC	X	X
3F	AAAT	X	X	3R	AAAC	X	X
4F	TTAA	X	X	4R	CCAA	X	X
5F	TAAA	X	X	5R	CAAA	X	X
6F	TAAT	X	X	6R	CCCA	X	X
7F	AACC	X	X	7R	AAGG	X	X
8F	ACCC	X		8R	AAAG	X	X
9F	AAAC	X		9R	GGAA	X	X
10F	CCAA	X		10R	GAAA	X	X
11F	CAAA	X		11R	GACA	X	X
12F	CCCA	X		12R	CTGA	X	
13F	AAGG	X					
14F	AAAG	X					
15F	GGAA	X					
16F	GAAA	X					

2.6. Chapter 7 materials and methods

2.6.1. Plant phenotyping

Seeds were initially sown in single cells (39 × 36 mm) to allow them to germinate before they were transferred to the growth pots (254 mm). Asparagine solution (0.1 M) was sprayed over the soil after sowing to ensure seed germination.

The seven genome edited lines were compared to a wildtype control. The lines were arranged following an incomplete block design, under non-radiating lights, with multiple plants in a single pot

which was large enough to ensure that the plants were not competing for resources and that their canopies did not compete for light. Due to some germination issues, the plants reached anthesis at very different times and were harvested at different times. The statistical analysis was performed by Andrew Mead (Rothamsted Research) using the GenStat statistical package (2018, 19th edition, © VSN International Ltd, Hemel Hempstead, UK).

The initial plan was to compare ten replicates of each genome edited line with twenty replicates of the wildtype control. There were eighteen pots, each to contain five plants, arranged in pairs under nine uniform light sources. Each pot would contain at least one wildtype control plant (with two pots containing two wildtype plants). Eight replicates of the genome edited lines were allocated following a doubly replicated balanced incomplete block design for seven treatments in fourteen blocks (pots) of size four, so four of the seven lines were in each pot, and a wildtype control was added to each of these pots. The remaining two replicates of each genome edited line and six replicates of the wildtype control were allocated to four extra blocks (pots) providing as close to balance as possible with two of these pots containing two wildtype controls. The original design was partially balanced for differences between the nine light sources; however, errors in the pairing of the pots removed this balance. Furthermore, one Line 99 plant unexpectedly died and was replaced with a Line 59 plant, further impacting the design structure, and meaning that the analysis required a restricted maximum likelihood (REML) approach for a linear mixed model. The plants were assessed for: height of tallest tiller, tiller and ear number, above ground biomass which was portioned into the stem and the ears, average grain weight and estimated thousand grain weight (measured from 50 seed weight).

The mean 50 seed weight for plant 15.2 (Line 99) was set to missing, to avoid incorrectly distorting the calculation of the adjusted mean value, as the plant produced no seeds. It was noted that the plant was particularly small, although not the smallest plant, and potentially unrepresentative of the growth response; however, the plant seemed to follow normal development, forming three ears (which contained no seed), and so the observed values were retained for the other variables. Another plant (plant 2.4; WT) showed incomplete ear filling, resulting in a zero top seed weight section, and this value was also set to missing.

2.6.2. Amino acid analyses in the T2 seed

Dried T2 seeds from the chosen lines were weighed and then ground to a fine powder under liquid nitrogen using a mortar and pestle. Pre-chilled solution A (500 mL methanol, 200 mL chloroform, 200 mL water) was added to the sample, using 1 mL of solution A for every 50 mg of tissue obtained, then the solution was vortexed and left on ice for 30 minutes. The sample was then vortexed and centrifuged at 14000 rpm for 2 minutes. The supernatant was extracted and placed on ice. The pellet was re-extracted with cold Solvent Mix B (500 mL methanol, 500 mL chloroform). The resuspended pellet was vortexed and then left on ice for a further 30 minutes. This was then vortexed again and

centrifuged at 14000 rpm for 2 minutes. The two supernatants were combined and then the aqueous and organic phases were separated by the addition of chilled water.

The two phases were placed in separate tubes and the samples were re-centrifuged. The aqueous phases were carefully transferred into fresh tubes and stored at -80 °C. The free amino acids were then analysed by Curtis Analytics (Rothamsted Research Campus, Harpenden, UK). The samples were derivatised using an EZfaast Amino Acid Analysis Kit (Phenomenex, Macclesfield, UK). Gas chromatography-mass spectrometry (GC-MS) analysis of the derivatised samples was carried out using an Agilent 6890 GC-5975-MS system (Agilent, Santa Clara, CA) in electron impact mode, as described previously (Curtis *et al*., 2018b; Elmore *et al*., 2005). The data were generated using the Agilent Chem Station Data analyses application.

2.6.3. Amino acid analyses in the T3 seed

The T3 amino acid analysis was also performed by Curtis Analytics but High Performance Liquid Chromatography (HPLC) was used due to a change of procedure at the analytical laboratory. The T3 grain was milled to wholemeal flour using a coffee grinder and 0.5 g of flour was used for each analysis. Free amino acids were extracted as described by Curtis *et al*. (2018b) (based on a method outlined by Noctor and Foyer in 1998), derivatised with o-phthalaldehyde (OPA) and analysed by HPLC. An HPLC Agilent 1100 (Agilent Technologies, US) was used, equipped with a Kinetexcolumn (2.6 µm XB-C18 100 A LC Column 150 × 4.6 mm; Phenomenex, UK) with an FLD detector (Agilent G1321D) to detect amino acids (using an excitation wavelength of 340 nm and emission wavelength of 455 nm).

The OPA reagent (15 µL) and the sample (10 µL) were mixed in the injection needle, and allowed to react for 1 minute, before injecting on to the column. Eluents were (A) 80 % 50 mM sodium acetate (pH 5.9) in 19 % methanol and 1 % tetrahydrofuran (THF) and (B) 80 % methanol in 20 % 50 mM sodium acetate (pH 5.9) at 0.8 mL/minute flow rate.

The column was heated to 30 °C and the total run time was 46 minutes, with the following run conditions: eluent A 100 %, 0 - 1 minutes; eluents A (90 %) and B (10 %), 6 - 11 minutes; eluents A (55 %) and B (45 %), 16 - 20 minutes; eluent B (100 %), 32 - 40 minutes; eluent A (100 %), 41 - 46 minutes. The data were obtained and analysed using Chemstation32 (v. B.04.03; Agilent Technologies).

2.6.4. *Cas9* integration

The integration of the *Cas9* gene was assessed by PCR amplification. The 2× ReddyMix PCR Master Mix (Thermo Fisher Scientific, Hemel Hempstead, UK) was used, following the manufacturer's instructions, with an annealing temperature of 57 °C and 100 ng/µL plant DNA. The primers were

UbiPro4 (5’-TTTAGCCCTGCCTTCATACG) and Cas9-SR1 (CACCTTCGCCATCTCGTTGC). The *Cas9*-containing plasmid (pRRes.486), which was originally used to transform the isolated embryos, was used as the control in the PCR. The PCR was performed on DNA extracted from leaf samples.

Chapter 3

Genomic Analysis of the Asparagine Synthetase Genes of Wheat and Other Cereals

Foreword

Parts of the work described in this chapter have been published in the following papers:

> Xu, H., Curtis, T.Y., Powers, S.J., Raffan, S., Gao, R., Huang, J., Heiner, M., Gilbert, D.R. and Halford, N.G., 2018. Genomic, biochemical, and modeling analyses of asparagine synthetases from Wheat. *Frontiers in Plant Science*, *8*, p.2237.
>
> Curtis, T.Y., Raffan, S., Wan, Y., King, R., Gonzalez-Uriarte, A. and Halford, N.G., 2019. Contrasting gene expression patterns in grain of high and low asparagine wheat genotypes in response to sulphur supply. *BMC Genomics*, *20*(1), p.628.
>
> Raffan, S. and Halford, N.G., 2021. Cereal asparagine synthetase genes. *Annals of Applied Biology*, doi: 10.1111/aab.12632.

3.1. Introduction

The development of crop varieties with reduced acrylamide-forming potential may enable the food industry to comply with regulations without costly changes to manufacturing lines or reduced product quality. This is discussed in detail in Chapter 1. In the case of wheat and other cereals, this means varieties with reduced and more consistent free asparagine concentration in the grain. There is, therefore, a need for greater knowledge and understanding of the genetic control of asparagine synthesis, accumulation and breakdown. A network of genes, enzymes and other factors involved in asparagine metabolism has been constructed (Curtis *et al*., 2018a), and at the centre of that network is asparagine synthetase, the enzyme that catalyses the transfer of an amino group from glutamine to aspartate to produce glutamate and asparagine.

3.1.1. Development of nucleotide sequencing technologies for reference genomes

With the advent of new and improved nucleotide sequencing technology, there is now a wealth of genomic information available, such as genetic maps and high-quality reference genomes. This has opened up new avenues in understanding the evolution and domestication of crops, become a useful tool in plant breeding and allowed for targeted crop improvement. Reference genomes, from genome sequence assemblies, underpin much of modern crop research. Historically, the large size, repeat-rich structure and polyploid nature of many crop genomes have made constructing these genome sequence assembles very difficult. However, the rapid improvements in genome sequencing technologies and bioinformatics in recent years have now made it feasible.

Early genome projects used bacterial artificial chromosomes (BACs) and Sanger sequencing. BACs were F-plasmid-based DNA constructs with an insert size typically of 150 - 350 kbp. This was a difficult and time-consuming process even for relatively small genomes, such as the human genome. Consequently, it was not attempted with many plant species and, when it was, it required a collaborative approach, with the focus on the genomes of key crops (rice (Sasaki *et al*., 2005), maize (Schnable *et al*., 2009), wheat (Choulet *et al*., 2014), and barley (Schulte *et al*., 2009), for example). The development of next generation sequencing (NGS) and computational technologies led to very high throughput, short read sequencing, with Illumina sequencing adopted in 2010 (Li *et al*., 2010; Li and Homer, 2010), and an increase in the number of plant genomes generated (e.g. amaranth (Clouse *et al*., 2016), squash (Sun *et al*., 2017), raspberry (VanBuren *et al*., 2016), carrot (Iorizzo *et al*., 2016), and 12 species of the Oryza genus (Jacquemin *et al*., 2013)).

NGS technologies made it feasible to sequence the whole genomes of the large-genome cereals, but initial attempts to generate contig assemblies using only short reads were not particularly successful (Chapman *et al*., 2015; International Wheat Genome Sequencing Consortium *et al*., 2014; Mayer *et*

al., 2012), with short-read sequencing mainly being used for linkage maps (Huang *et al*., 2009; Mascher *et al*., 2013), physical maps (van Oeveren *et al*., 2011), and chromosome-specific sequences (Mayer *et al*., 2009) in these species. The generation of the short reads was costly and the bioinformatics assembly methods were lacking, making it an unattractive proposition. The first high-quality cereal genome reference sequence, or reference sequence for any repeat-rich polyploid plant genome for that matter, was an emmer wheat (*Triticum dicoccoides*; also known as *Triticum turgidum* subsp. *dicoccoides*) assembly that was constructed from Illumina sequencing data (Avni *et al*., 2017). Multiple paired-end reads were used, where the library fragments are read from both ends, alongside mate-pair libraries, which provides longer reads, and chromosome-conformation capture, which crosslinks the DNA to give information on the chromatin structure. This was possible due to improved methods in library preparation, which allowed for uniform genome representation (Aird *et al*., 2011), improved Illumina technology and greatly decreased cost of sequencing, underpinned by developments in the scaffold construction software. A scaffold is constructed from contigs (long continuous stretches of sequence) that have been assembled together.

The development of long read (> 10 kbp) sequencing technologies, such as those from PacBio or Oxford Nanopore, reduced the difficulty of genome assemblies; however, this came at a cost of reduced sequence accuracy (Koren *et al*., 2012). Some plant genomes have been reassembled using these methods (e.g. maize (Jiao *et al*., 2017); tomato wild relative (Schmidt *et al*., 2017)); however, correction with short reads may still be required. A combination of long and short read sequencing data was used to reconstruct the wheat (*Triticum aestivum*) and goat grass (*Aegilops tauschii*) genomes (Zimin *et al*., 2017; Luo *et al*., 2017b), and this proved to be an effective method to provide highly accurate assemblies. However, this was only possible because of computational developments that enabled the immense computational requirements to be met (Zimin *et al*., 2017).

Sequencing technology has continued to develop considerably in the last few years. Contigs of megabase-scale contiguity are now possible and many methods now exist to correct sequence assemblies (Lam *et al*., 2012; Zheng *et al*., 2016a; Lieberman-Aiden *et al*., 2009). This is important because complete, contiguous and annotated reference sequences are critical in supporting research.

3.1.2. Use of the wheat reference sequence in research and breeding

Genomic tools are particularly important in plant breeding and crop improvement. Plant breeding is conducted on the genome-wide scale, with each cross bringing large changes in the genome-wide networks that underpin complex traits of interest, such as yield, disease protection and quality. Development can be limited by lack of knowledge of the underlying molecular biology, especially when phenotypic traits, the basis for most selection, rely on the expression of multiple loci or are affected by environmental factors, and by the interactions of genetic and environmental factors (G × E). Nevertheless, the identification of potential candidate genes can allow for accurate selection and

trait development. The discovery of novel single nucleotide polymorphisms (SNPs) is critical in marker-assisted selection programmes and genomic resources are invaluable in favourable trait detection. Access to a fully annotated reference sequence has increased the understanding of important traits in many plant and animal species (reviewed by Hickey *et al*., 2017).

Wheat has been particularly difficult to produce a high-quality reference sequence for, due to the large allopolyploid nature of its genome, with the wheat genome being over five times larger than the human genome and highly repetitive (International Wheat Genome Sequencing Consortium, 2014; Appels *et al*., 2018; Alaux *et al*., 2018), containing over 85 % repetitive DNA. The International Wheat Genome Sequencing Consortium has been working on wheat since 2005, recently generating an annotated reference genome (IWGSC RefSeq v1.0) with all the chromosomes structurally organised (Appels *et al*., 2018). This provides the foundation for the improved understanding of wheat molecular biology, allowing for new DNA marker platforms to be developed, as well as targeted breeding technologies, such as genome editing (reviewed in Chapter 1).

3.1.3. The importance of reference sequences in genome editing

CRISPR/Cas9 is an effective tool for genome manipulation and for generating mutants. However, it requires precise targeting through the gRNA guide sequence and the presence of a PAM sequence, in this case NGG (Cong *et al*., 2013). The genomic sequence of the target gene is required, with a knowledge of the intron/exon structure of the gene so that exons can be targeted, rather than introns where the edits are likely to have little to no effect, making sure that the target site does not lie over an intron. Furthermore, if a perfect match is to be designed to the target site, the sequence of the variety to be used is preferable to ensure there are no unexpected single nucleotide polymorphisms. Since cv. Cadenza is the preferred variety for biolistic transformation of *Triticum aestivum* at Rothamsted, with one of the highest transformation efficiencies (Sparks and Jones, 2009), the genes needed to be investigated using a cv. Cadenza reference sequence which was assembled by the Grassroots Infrastructure project (Bian, *et al*., 2017; Gardiner *et al*., 2019; Krasileva *et al*., 2017).

3.1.4. Expression analysis of the *TaASN* genes

Expression analysis is important to understand the roles of different genes in different tissues and the complex regulatory networks involved in specific biology processes.

Free asparagine accumulates in many plant tissues in response to a range of abiotic and biotic stresses, as well as during normal physiological processes such as seed germination (Lea *et al*., 2007). It was known that the *TaASN* genes were differentially expressed in plant tissues (Gao *et al*., 2016) and that free asparagine accumulates in wheat grain in response to sulphur deficiency (Muttucumaru *et al*., 2006; Curtis *et al*., 2009; Curtis *et al*., 2010; Curtis *et al*., 2018b; Granvogl *et al*., 2007) and poor

disease control (Curtis *et al*., 2016). There are also substantial differences in the free asparagine concentration of grain from different wheat varieties and genotypes (Curtis *et al*., 2018b).

In the present study, an RNA-seq dataset was analysed to investigate the expression of the different wheat asparagine synthetase genes in embryo and endosperm. The dataset had been produced by Dr Tanya Curtis (Rothamsted Research) to compare two wheat genotypes, variety Spark and a doubled haploid line, SR3, from a Spark × Rialto mapping population (Snape *et al*., 2007), grown under conditions of sulphur sufficiency and deficiency. Developing seeds had been sampled at 14 and 21 days post anthesis (dpa). The two varieties have differing free asparagine concentrations, with SR3 showing a lower concentration of free asparagine in the grain than Spark (Curtis *et al*., 2009).

3.1.5. The *ASN* gene families

The wealth of genomic information available and the high quality of the genome assemblies allow for genome mining and the investigation of gene families *in silico*. In fact, the *in-silico* characterisation of gene families, often complimented by expression data, has become widespread and routine when investigating the role of a gene (for example, *RBOH* genes in fruit trees (Cheng *et al*., 2019); *BES1* in soybean (Li *et al*., 2019); *DGAT* in oil palm (Rosli *et al*., 2018)). Negative phenotypes have been seen in *asn* mutants in other plant species, meaning that characterising the *ASN* genes, in *Triticum aestivum* specifically, but across a range of cereal species, was particularly important.

In potato, for example, a genetically modified variety was developed with reduced asparagine synthetase gene expression in the tubers (Chawla *et al*., 2012). Potato has two *ASN* genes, *StASN1* and *StASN2* (Rommens *et al*., 2008). The simultaneous silencing of both genes in the tuber through RNA interference produced potatoes with cracked and small tubers when placed in the field (Chawla *et al*., 2012), although no phenotypic effects were observed in the glasshouse trials (Rommens *et al*., 2008). Tuber specific RNAi was picked because it was already known that *ASN* genes were important in photorespiration (Ta *et al*., 1985) in peas and ammonium detoxification (Wong *et al*., 2004) in the leaves of Arabidopsis.

Rice also has two asparagine synthetase genes (Watanabe *et al*., 1996; Nakano *et al*., 2000; Sakai *et al*., 2013), *OsASN1* and *OsASN2*. The genes show differing expression patterns, with *OsASN1* expression mainly confined to the roots and *OsASN2* expressed in the leaves (Ohashi *et al*., 2015). Luo *et al*. (2019) showed that *OsASN1* is involved in rice development, including tiller outgrowth. The functional knock-out of the *OsASN1* gene resulted in decreased asparagine concentrations throughout the plant, reduced growth and a reduction in tiller bud development. The expression of many other genes in the asparagine metabolism pathway was affected. Notably, *OsASN2* expression was significantly higher in the *osasn1* mutant, presumably a result of the low levels of asparagine and

a compensation mechanism in the plant. The expression of other asparagine metabolism genes, including genes involved in glutamine and glutamate transformation, declined in the mutant plants.

The cDNA sequences had been discovered for *TaASN1*, *TaASN2* and *TaASN3* (Genbank BT009245, BT009049 and AK333183, respectively) prior to this study, but little was known about the *TaASN4* sequence (Gao *et al*., 2016). Gao *et al*. (2016) examined the expression of *TaASN1*, *TaASN2* and *TaASN3* and showed *TaASN2* to be the most highly expressed in the grain, making *TaASN2* the logical target for genetic interventions. *TaASN1*, however, was shown to be the most environmentally responsive (Gao *et al*., 2016; Wang *et al*., 2005; Curci *et al*., 2018). Very little was known about the expression of *TaASN4* (Gao *et al*., 2016).

3.1.6. Chapter Aims

This Chapter describes a genomic analysis of the asparagine synthetase gene families not only of bread wheat (*Triticum aestivum*) and its close relatives but also other cereals: barley (*Hordeum vulgare*), rye (*Secale cereale*), brachypodium (*Brachypodium dystachion*), maize (*Zea mays*), sorghum (*Sorghum bicolor*) and rice (*Oryza sativa*). The aim was to determine how the asparagine synthetase gene family was organised in the different cereal species and how it had evolved. The results may also be useful in explaining the differences in free asparagine levels in the different species as data on that become available. They will also inform strategies for genetic interventions to reduce free asparagine levels in the different species, including how 'transferable' the technology used in this project is likely to be.

The RNA-seq analysis aimed to confirm that *TaASN2* is the appropriate target for the genetic interventions leading to grain-specific reduction in free asparagine concentration in wheat, and whether additional targets would be needed for the effective reduction of asparagine accumulation. The analysis was also required to properly interpret any results gained from the genome editing interventions. If there was no phenotype, then the expression analysis would be needed to explain why this was, in the context of the underlying molecular network.

3.2. Results

3.2.1. Investigating the *TaASN* genes in cv. Chinese Spring and cv. Cadenza

The *ASN* gene family was investigated in both wheat (*Triticum aestivum*) cv. Chinese Spring (NR Gene v4 and IWGSC RefSeq v1.0 (Alaux, *et al*., 2018) and cv. Cadenza (NR Gene v4). The presence of five *ASNs* (Gao *et al*., 2016), *TaASN1-4*, with a double copy of *TaASN3*, per genome was confirmed, with a total of 14 *TaASNs* in Chinese Spring (NR Gene v4) and 15 in Cadenza.

The cDNA sequences of *TaASN2* (Genbank BT009049), *TaASN1* (Genbank BT009245) and *TaASN3* (Genbank AK333183) were used to BLAST the NR Gene v4 and IWGSC RefSeq v1.0 Chinese Spring reference databases to identify the total number of *TaASN* genes and assess their similarity, as well as their corresponding gene positions.

Table 3.1. Current Gene IDs in the Ensembl database for the wheat (*Triticum aestivum*) cv. Chinese Spring asparagine synthetase genes (PLANT_T. aestivum_NRgene_v0.4_scaf).

Gene name		Gene ID in Emsembl	Position on forward strand
TaASN1	A	TraesCS5A02G153900	Chromosome 5A: 331,398,338-331,403,028
	B	TraesCS5B02G152600	Chromosome 5B: 281,836,802-281,840,213
	D	TraesCS5D02G159100	Chromosome 5D: 247,610,358-247,613,753
TaASN2	A	TraesCS3A02G077100	Chromosome 3A: 47,847,824-47,852,130
	D	TraesCS3D02G077300	Chromosome 3D: 37,774,235-37,785,576
TaASN3.1	A	TraesCS1A02G382800	Chromosome 1A: 553,535,726-553,542,082
	B	TraesCS1B02G408200	Chromosome 1B: 635,920,024-635,926,285
	D	TraesCS1D02G390500	Chromosome 1D: 461,902,232-461,908,026
TaASN3.2	A	TraesCS1A02G422100	Chromosome 1A: 577,840,627-577,846,994
	B	TraesCS1B02G453600	Chromosome 1B: 669,754,736-669,761,058
	D	TraesCS1D02G430300	Chromosome 1D: 481,384,495-481,389,740
TaASN4	A	TraesCS4A02G109900	Chromosome 4A: 133,453,278-133,457,252
	B	TraesCS4B02G194400	Chromosome 4B: 417,737,785-417,741,607
	D	TraesCS4D02G195100	Chromosome 4D: 338,823,037-338,827,058

The chromosomal positions and the Ensembl gene IDs are given in Table 3.1. *TaASN1*, *TaASN2*, and *TaASN4* were all found to be single copy genes, located on chromosomes 5, 3, and 4, respectively, of each genome (A, B, and D), except that *TaASN2* was not present in the B genome of Chinese Spring. Analysis of the Cadenza genome showed that not all wheat varieties lack a *TaASN2* gene on chromosome 3B as Cadenza possesses all three homeologues of the *TaASN2* gene. In the case of *TaASN3*, there were two copies on chromosome 1 of each genome, indicating a recent gene duplication event, and these have been annotated as *TaASN3.1* and *TaASN3.2* (Xu *et al*., 2018). In other words, cv. Cadenza was shown to contain five *ASN* genes per genome, *TaASN1*, *TaASN2*, *TaASN3.1*, *TaASN3.2* and *TaASN4*, but the cv. Chinese Spring B genome had only four because it lacked *TaASN2*.

The structures of the genes are shown (Figure 3.1), illustrating the considerable divergence of intron/exon patterns between the different genes, but conservation of structure within each group of homeologues.

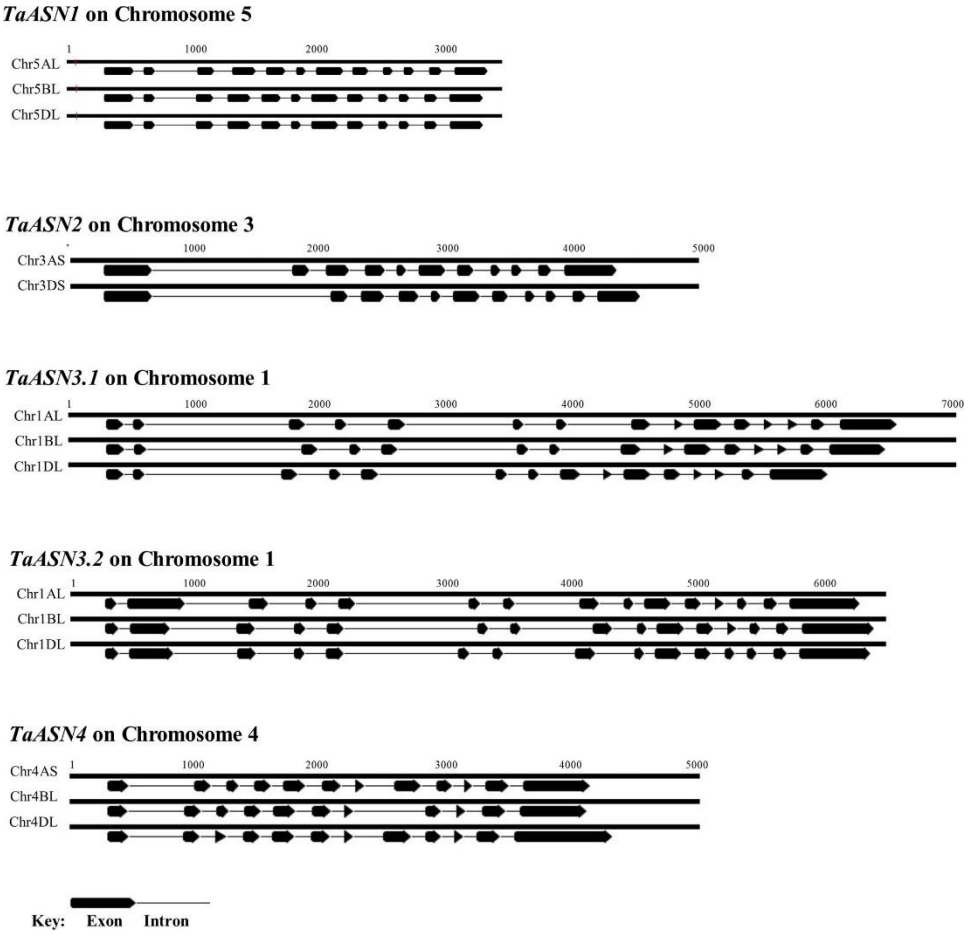


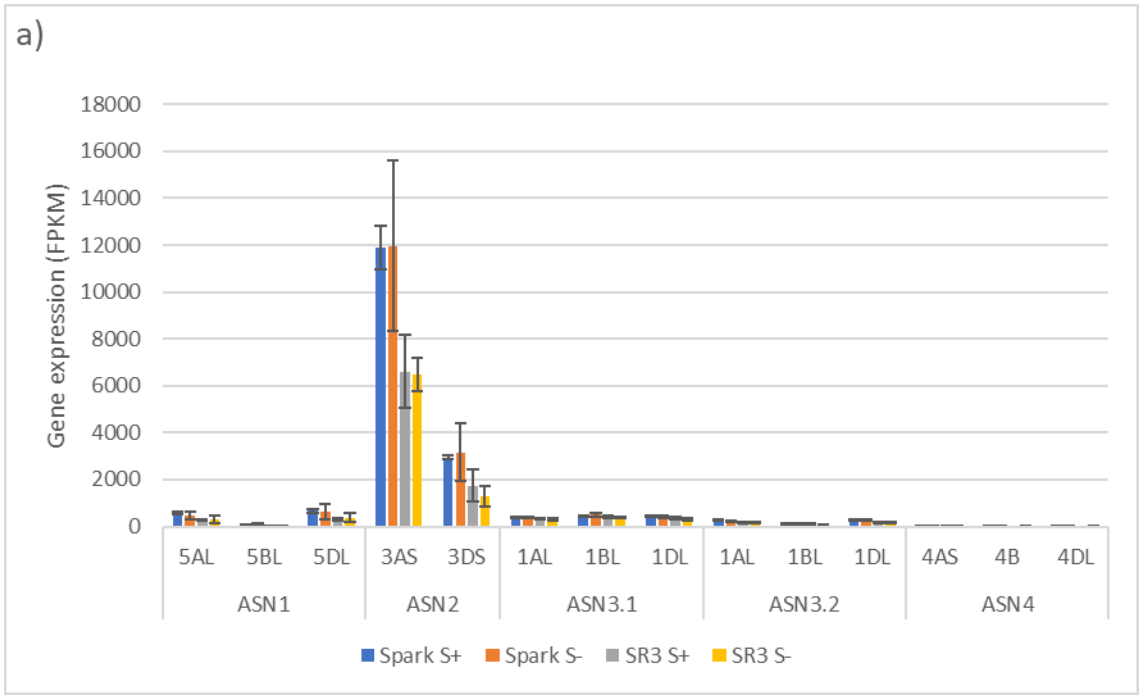
Figure 3.1. Diagrammatic representation of the gene structures of *TaASN1*, *TaASN2*, *TaASN3.1*, *TaASN3.2*, and *TaASN4*, based on the wheat (*Triticum aestivum*) cv. Chinese Spring reference sequence (PLANT_*T. aestivum*_NRgene_v0.4_scaf).

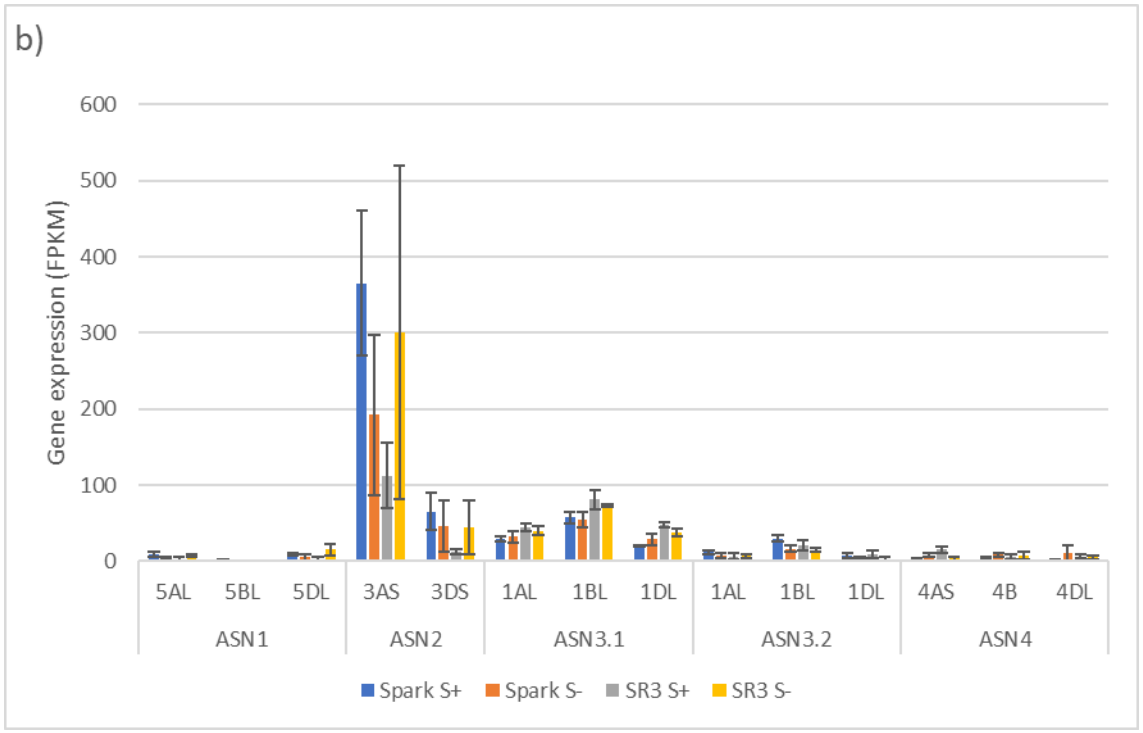
The three *TaASN1* homeologues, on chromosomes 5A, B, and D, are the shortest at approximately 3 kbp from the ATG translation start codon to the translation stop codon, including 12 exons. The two *TaASN2* homeologues are approximately 4 kbp in length, with 11 exons, and the three *TaASN3.1* and *TaASN3.2* homeologues just over and just under 6 kbp, respectively, making them the longest group. The two *TaASN3* genes share a similar intron/exon pattern, with 15 exons. The three *TaASN4* homeologues are just under 4 kbp in length, with 12 exons, except that the *TaASN4* gene on chromosome 4B lacks exon 8. Clearly, this deletion may affect the activity of the enzyme encoded by the gene. The relatively simple structure of the gene family meant that genetic interventions to reduce free asparagine accumulation and thereby acrylamide-forming potential in wheat grain were more likely to be successful.

An additional gene, with all three homeologues, was identified when the INSDC Assembly GCA_900519105.1 (Ensembl) was searched for genes containing an asparagine synthetase domain using the InterPro asparagine synthetase domain IPR001962. The homeologues *TraesCS5A02G081800*, *TraesCS5B02G084600* and *TraesCS5D02G090800*, were identified on chromosomes 5A, 5B and 5D, respectively; however, the gene was shown not to be an asparagine synthetase and the function of the protein encoded by this gene is not known (section 3.5.2).

3.2.2. RNA-Seq analysis

The study involved the analysis of a dataset that was already available (Curtis *et al*., 2019). Two genotypes, variety Spark and doubled haploid line SR3 from a Spark × Rialto mapping population, had been grown under glass with sulphur supplied or not supplied. Embryo and endosperm tissue had been sampled at 14 and 21 days post anthesis (dpa) and gene expression analysed in the two tissues by RNA-seq. The analysis was performed in collaboration with Robert King and Asier Gonzalez-Uriarte of the Computational and Analytical Sciences Department at Rothamsted Research. Consistent with previous research (Gao *et al*., 2016), higher levels of total asparagine synthetase gene expression were seen in the embryo than in the endosperm (> 10-fold difference) (Figure 3.2). *TaASN2* was confirmed to be the most highly expressed *TaASN* gene in the grain, in both the embryo and the endosperm, in both sulphur conditions and at both timepoints. In fact, the *TaASN2* gene accounted for over 2/3 of the total *TaASN* gene expression, despite the lack of the B genome allele, which is absent in both Spark and SR3 (all the reads could be assigned to the A or D genome homeologue), as is seen in Chinese Spring (Xu *et al*., 2018). The A genome gene was shown to be the more highly expressed homeologue of *TaASN2* (> 3-fold difference compared to the D genome gene) in both genotypes, at both timepoints and for both sulphur conditions, with the A genome alone accounting for more than half the total *TaASN* gene expression. This led to the hypothesis that targeting only the A genome homeologue may be enough to reduce asparagine accumulation in the grain whilst reducing the risk of negative phenotypic effects that might be caused by a full-gene knock-out (Luo *et al*., 2019; Chawla *et al*., 2012).


a)
Gene expression (FPKM)
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
5AL
5BL
5DL
3AS
3DS
1AL
1BL
1DL
1AL
1BL
1DL
4AS
4B
4DL
ASN1
ASN2
ASN3.1
ASN3.2
ASN4
Spark S+
Spark S-
SR3 S+
SR3 S-


b)
Gene expression (FPKM)
600
500
400
300
200
100
0
5AL
5BL
5DL
3AS
3DS
1AL
1BL
1DL
1AL
1BL
1DL
4AS
4B
4DL
ASN1
ASN2
ASN3.1
ASN3.2
ASN4
Spark S+
Spark S-
SR3 S+
SR3 S-

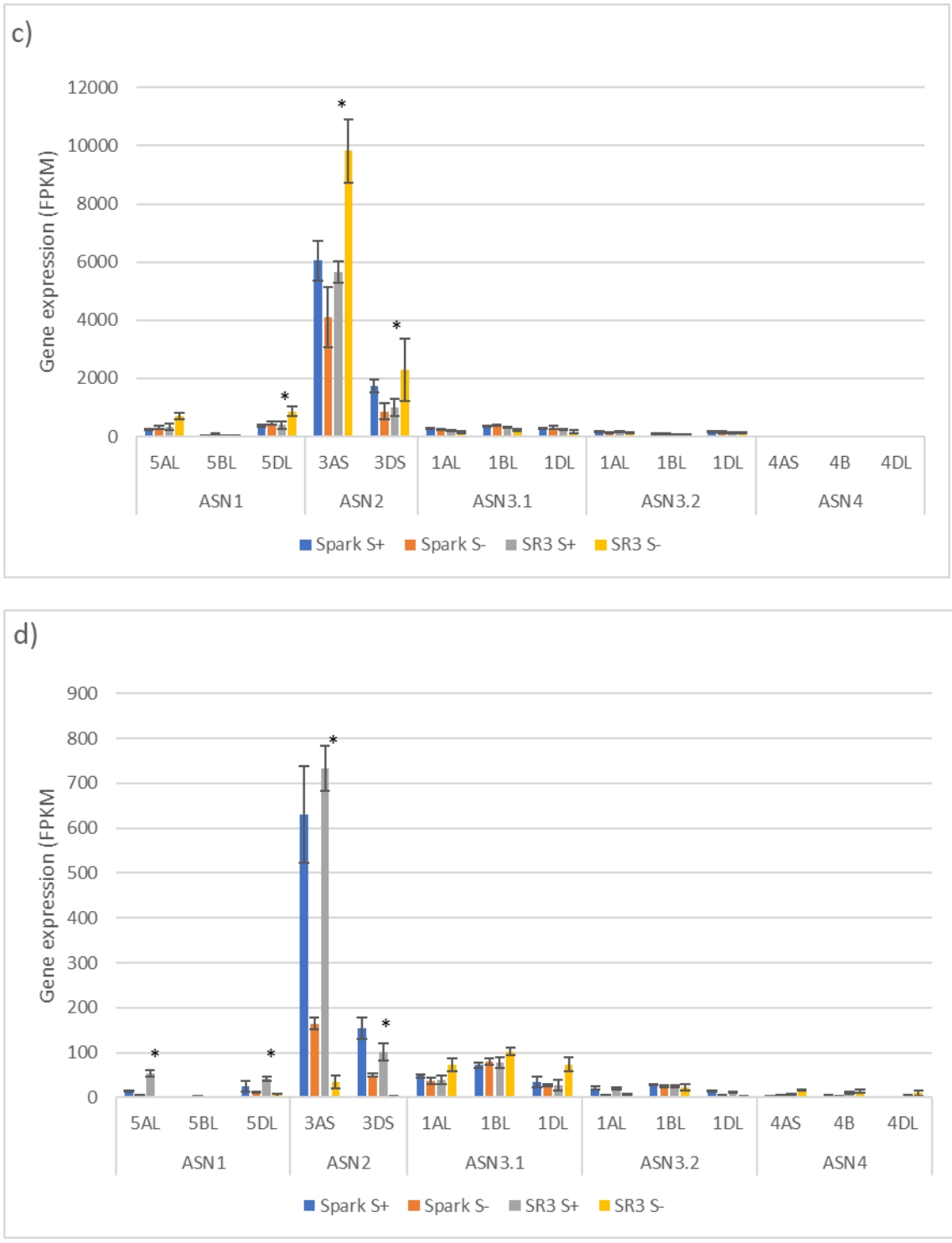


Figure 3.2. Expression levels (FPKM) of the *TaASN* genes in the embryo and endosperm of the grain of genotypes, Spark and SR3, grown with or without sulphur. a) Embryo, 14 dpa. b) Endosperm, 14 dpa. c) Embryo, 21 dpa. d) Endosperm, 21 dpa. Significant differences are seen in the *TaASN1* 5D homeologue ($p = 0.0478$) and both *TaASN2* homeologues ($p = 0.038$ and 0.047 for the A and D respectively) in SR3 in the embryo at 21 dpa; and the two *TaASN2* homeologues ($p < 0.001$ for both) and the A and D homeologues of *TaASN1* ($p < 0.001$ and $p = 0.0012$ respectively) in the endosperm of SR3 at 21dpa.

The dataset also revealed differences in the expression of the *TaASN1* homeologues, with the B genome gene showing lower expression levels than the other homeologues. *TaASN3.1* showed higher levels of expression than *TaASN3.2*, and it was the second most highly expressed *TaASN* gene in both tissues at 21 days. *TaASN4* showed very low levels of expression in all the samples.

Variety Spark has been shown to have a higher concentration of free asparagine in the grain than SR3 (the concentration in SR3 being lower than either parent of the mapping population, Spark or Rialto) (Curtis *et al*., 2009) and this might be explained by the higher *TaASN2* gene expression in Spark at 14 dpa. The response to sulphur deficiency was complex, with neither genotype showing a significant sulphur response at 14 dpa. However, a response to sulphur deficiency was seen in SR3 at 21 dpa, with the D genome *TaASN1* increasing in expression (and A genome in the embryo only), and both *TaASN2* homeologues showing increased expression in both tissues.

The analysis was widened to focus on other genes involved in asparagine synthesis, turnover and accumulation (Curtis *et al*., 2018), with the aim of understanding the role of sulphur in asparagine metabolism and the differences between the genotypes. Under sulphur sufficient conditions, the genes encoding enzymes of nitrogen assimilation were shown to be more highly expressed in Spark than SR3. In the embryo of SR3, glutamine synthetase gene expression increased in response to sulphur deficiency and asparaginase gene expression decreased. The genes encoding SnRK1 and GCN2, regulatory protein kinases which regulate asparagine synthetase gene expression (Baena-González *et al*., 2007; Byrne *et al*., 2012), were also shown to respond to sulphur deficiency.

This highlights other potential targets for genetic interventions into asparagine metabolism; however, *TaASN2* was confirmed to be the best target for genome editing. The differential expression of the homeologues allows for the targeting of the A genome on its own, to knockdown rather than knockout *TaASN2*.

3.2.3. Phylogenetic analysis of cereal asparagine synthetase genes

3.2.3.1. The Triticeae: Bread wheat (*Triticum aestivum*)

Wheat is a name assigned to a number of species within the Triticeae tribe of the Pooideae subfamily of the family Poaceae. As discussed in Section 3.3.1, the *TaASN* gene family of bread wheat comprises five genes per genome (Gao *et al*., 2016; Xu *et al*., 2018) (Table 3.1). *TaASN3.1* and *TaASN3.2* were confirmed as paralogues resulting from a relatively recent gene duplication (Section 3.3.1) (Xu *et al*., 2018).

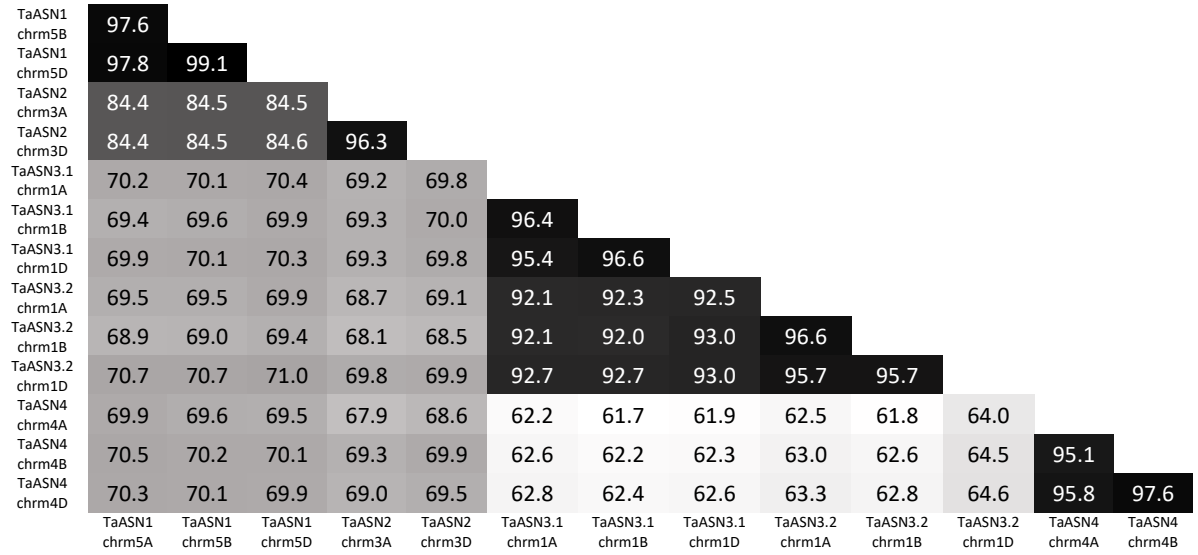


Figure 3.3. Similarity matrix of the *TaASN* genes calculated from a nucleotide alignment of the genes generated using Geneious (Kearse *et al*., 2012). The values presented are the percentage of bases that are identical, expressed to one decimal place.

The nucleotide sequence data for *TaASN* genes (cv. Chinese Spring; International Wheat Genome Sequencing Consortium (IWGSC), 2018, http://plants.ensembl.org/index.html) was used to generate a similarity matrix (Figure 3.3). The *TaASN* genes with the highest similarity were, unsurprisingly, *TaASN3.1* and *TaASN3.2* (92.1 – 93.0 % nucleotide sequence identity). A high level of similarity was also seen between *TaASN1* and *TaASN2* (84.4 - 84.6 % identity). Within the genes themselves, *TaASN1* showed the highest similarity between the homeologues, at 97.6 - 99.1 % identity, while *TaASN2* showed 96.3 % between the A and D genome homeologues. The *TaASN3.1* homeologues shared 95.4 - 96.6 % identity, with the *TaASN3.2* homeologues sharing 95.7 – 96.6 % and *TaASN4* sharing 95.1 - 97.6 % identity (Figure 3.3).

A phylogenetic tree was constructed from a similarity matrix generated for all the *ASN* genes in the Triticeae species (Figure 3.4; matrix in supplementary file). This tree shows the assignment of the five *TaASN* genes to the four *ASN* groups. The analysis confirmed the lack of the B genome homeologue of *TaASN2* in cv. Chinese Spring (Section 3.3.1; Xu *et al*., 2018). This has been observed in cv. Spark and SR3 (Section 3.3.2; Curtis *et al*., 2019), suggesting that the B genome deletion of *TaASN2* may be widespread in modern wheat cultivars; however, the B genome homeologue has been shown to be present in the cv. Cadenza genome (Chapter 4; genome data available from EnsemblPlants) so the deletion is not universal. The presence of this deletion in modern wheat cultivars, and its effect, are being followed up in another PhD project.

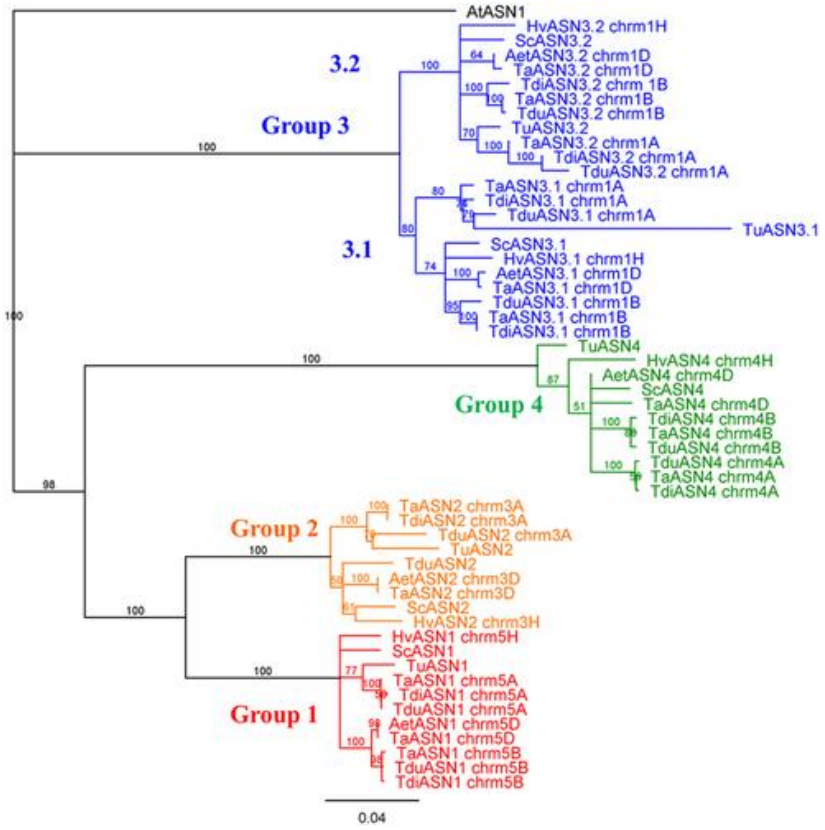


Figure 3.4. Phylogenetic tree of the asparagine synthetase genes in Triticeae species generated from a similarity matrix, with the Arabidopsis (*Arabidopsis thaliana*) gene, *AtASN1*, used as an outgroup to anchor the tree. The species included are bread wheat (*Triticum aestivum*; genes shown with a Ta prefix), emmer wheat (*Triticum dicoccoides*; Tdi prefix), durum (pasta) wheat (*Triticum durum*; Tdu prefix), einkorn wheat (*Triticum urartu*; Tu prefix), Tausch's goatgrass (*Aegilops tauschii*; Aet prefix), barley (*Hordeum vulgare*; Hv prefix) and rye (*Secale cereale*; Sc prefix). The branch labels show the consensus support in the clade from bootstrapping, as a percentage. The scale bar represents the length of the branches, expressed in units of substitutions per site of the sequence alignment.

3.2.3.2. The Triticea: Emmer wheat (*Triticum dicoccoides*)

Emmer or hulled wheat (*Triticum dicoccoides*, also known as *Triticum turgidum* subsp. *Dicoccoides*), along with einkorn wheat (*Triticum urartu*), was widely cultivated in the ancient world. It is a tetraploid wheat (genomes AABB) and is believed to be the wild progenitor of cultivated pasta wheat (*Triticum durum*) and *Triticum dicoccum* (also known as *Triticum turgidum* subsp. *Durum* and *Triticum turgidum* subsp. *Dicoccum*). It is also believed to be one of the progenitors of modern bread wheat (AABBDD), along with *Aegilops tauschii* (DD) (Salamini *et al*., 2002; Matsuoka, 2011). The genome assembly is based on the accession Zavitan, as genetic data were available for this genotype (WEWSeq v.1.0, INSDC Assembly).

Nine *ASN* genes were identified in *Triticum dicoccoides* (Table 3.2). Two genes, TRIDC5AG025640 and TRIDC5BG026790 (Ensembl reference numbers), were located on chromosomes 5A and 5B, respectively. Only one gene was located on chromosome 3, TRIDC3AG009140 on chromosome 3A. No B genome homeologue of this gene was identified. Two genes were located on chromosome 4, one of chromosome 4A, TRIDC4AG015200, and a homeologue on chromosome 4B, TRIDC4BG033760. Matching what was seen in bread wheat (Section 3.3.1; Xu *et al*, 2018), two genes were located on chromosome 1A, TRIDC1AG056280 and TRIDC1AG062090, and two genes were located on chromosome 1B, TRIDC1BG064720 and TRIDC1BG071210.

Table 3.2. Gene names (Raffan and Halford, 2021), Ensembl gene identifiers and wheat orthologues of the nine *ASN* genes identified in *Triticum dicoccoides.*

Gene name	Gene ID	Wheat orthologue	
TdiASN1	TRIDC5AG025640	*TaASN1*	A
	TRIDC5BG026790		B
TdiASN2	TRIDC3AG009140	*TaASN2*	A
TdiASN3.1	TRIDC1AG056280	*TaASN3.1*	A
	TRIDC1BG064720		B
TdiASN3.2	TRIDC1AG062090	*TaASN3.2*	A
	TRIDC1BG071210		B
TdiASN4	TRIDC4AG015200	*TaASN4*	A
	TRIDC4BG033760		B

The *TdiASN* genes show the same grouping as the *TaASN* genes (Figure 3.4), and were named according to the groups they fell into. The genes on chromosome 5 were identified as *TdiASN1s*, the gene on chromosome 3A was designated a *TdiASN2*, and the genes on chromosome 4 were designated *TdiASN4* genes (Raffan and Halford, 2021; Table 3.2). The genes on chromosome 1 were identified based on their similarity to the *TaASN3* genes, with TRIDC1AG056280 and TRIDC1BG064720 as *TdiASN3.1* and TRIDC1AG062090 and TRIDC1BG071210 as *TdiASN3.2*. Only one *TdiASN2* gene was identified, with the B chromosome homeologue missing in the database. As stated above, the B genome version of *TaASN2* is missing in some bread wheat varieties but present in others. Since *Triticum dicoccoides* is regarded as the A and B genome progenitor of bread wheat, it is somewhat surprising that a B genome *TdiASN2* could not be identified.

An alignment matrix was generated for the *TaASN* and *TdiASN* genes (Figure 3.5). This highlighted the close evolutionary history between the species, with 100 % identity shown between *TdiASN2* and the chromosome 3A homeologue of *TaASN2*. The A genome homeologues of *TdiASN1* and *TaASN1* shared 99.5 % identity, with the B genome homeologues sharing 99.7 %. The A and B *TdiASN1* genes shared 97.3 % and 99.2 % identity, respectively, with the chromosome 5D *TaASN1*, which is comparable to what is seen between the *TaASN1s* (97.6 - 99.1 %). The *TdiASN1* A and B genome

homeologues share 83.8 – 84.1 % identity to the *TaASN2* genes, and the *TdiASN2* gene shows 84.3 – 84.4 % identity to the *TaASN1* genes.

The comparison of the *TdiASN3* and *TaASN3* genes showed 99.5 % and 100 % identity for the A and B genome genes, respectively, with the *TdiASN3.1* genes showing lower similarity to the *TaASN3.1* D homeologue (95.4 - 96.6 %). The A and B genome *ASN3.2* genes showed 99.6 % and 98.1 % identity, respectively, when compared to their *TdiASN* counterparts, and 95.4 % and 95.2 % when compared to the D genome *TaASN3.2*. The *TdiASN4* genes showed 99.2 % (A genome) and 99.6 % (B genome) to their respective *TaASN4* counterparts, and 97.3 – 97.7 % identity to the D genome *TaASN4* homeologue.

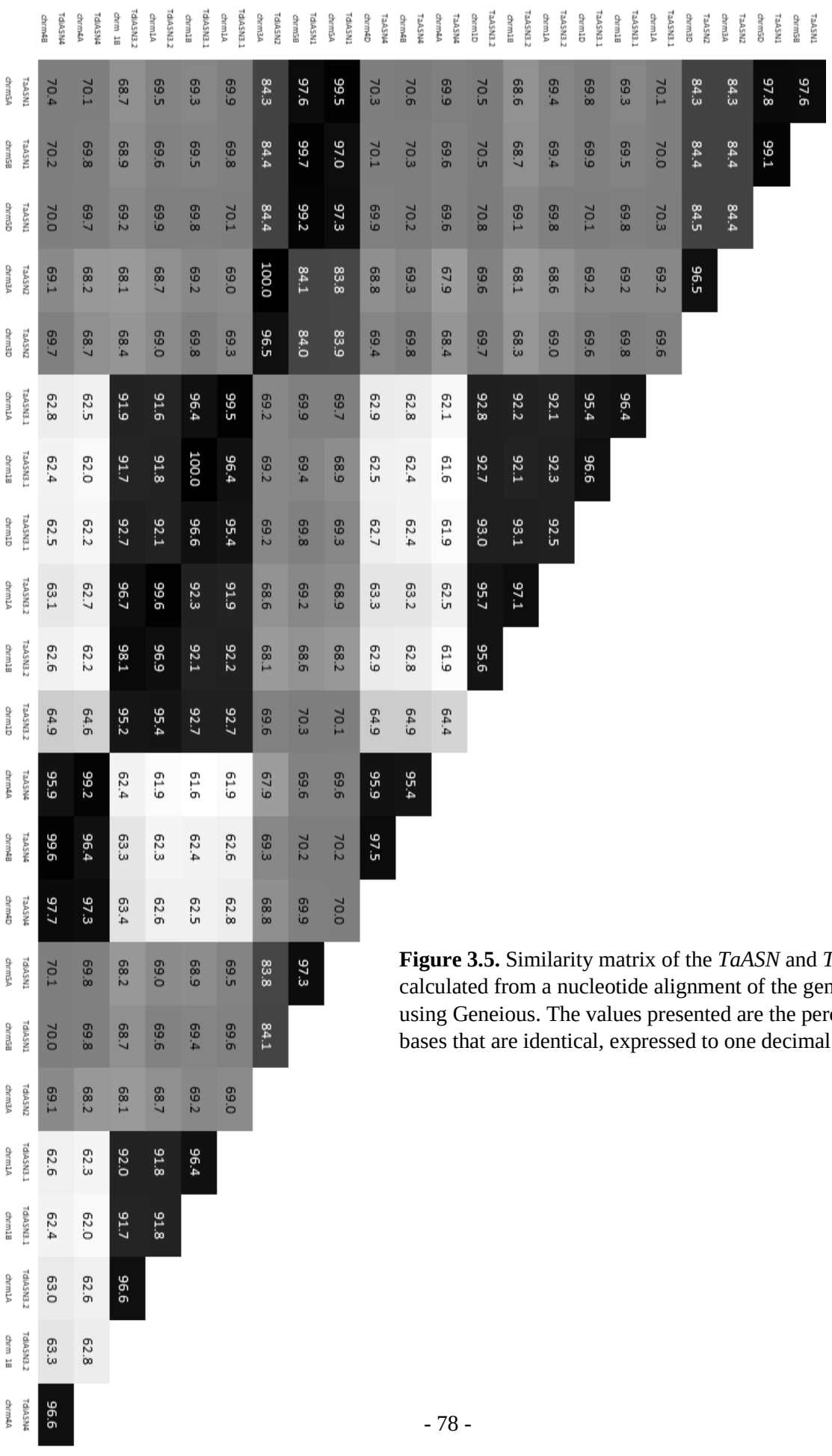


Figure 3.5. Similarity matrix of the *TaASN* and *TdiASN* genes calculated from a nucleotide alignment of the genes generated using Geneious. The values presented are the percentage of bases that are identical, expressed to one decimal place.

3.2.3.3. The Triticea: Pasta wheat (*Triticum durum*)

Cultivated pasta wheat (*Triticum durum*; also known as *Triticum turgidum* subsp. *Durum*) is considered to be the descendant of emmer wheat and is also a tetraploid (AABB) species (Peng *et al*., 2011). The reference sequence is based on the variety Svevo (Svevo.v1, INSDC Assembly GCA_900231445; Maccaferri *et al*., 2019), and the Svevo × Zavitan genetic map was used to order and orient the scaffolds.

Table 3.3. Gene names (Raffan and Halford, 2021), Ensembl gene identifiers and wheat orthologues of the ten *ASN* genes identified in *Triticum durum.*

Gene Name	Gene ID	Wheat orthologue	
TduASN1	TRITD5Av1G114980	*TaASN1*	A
TduASN1	TRITD5Bv1G097460		B
TduASN2	TRITD3Av1G021350	*TaASN2*	A
TduASN2	TRITD0Uv1G024260		B
TduASN3.1	TRITD1Av1G215640	*TaASN3.1*	A
TduASN3.1	TRITD1Bv1G207110		B
TduASN3.2	TRITD1Av1G222870	*TaASN3.2*	A
TduASN3.2	TRITD1Bv1G220260		B
TduASN4	TRITD4Av1G053230	*TaASN4*	A
TduASN4	TRITD4Bv1G119970		B

Two *ASN* genes were previously characterised in *Triticum durum* (Curci *et al*., 2018), whereas this analysis identified ten *ASN* genes (Table 3.3). Two genes were identified on chromosome 5, TRITD5Av1G114980 and TRITD5Bv1G097460 (Ensembl reference numbers), which were located on chromosomes 5A and 5B, respectively. These genes were designated as *TduASN1* genes. Two genes were located on chromosome 4, one on chromosome 4A, TRITD4Av1G053230, and a homeologue on chromosome 4B, TRITD4Bv1G119970. These genes were called *TduASN4*. Two genes were located on chromosome 1A, TRITD1Av1G215640 and TRITD1Av1G222870, and two genes were located on chromosome 1B, TRITD1Bv1G207110 and TRITD1Bv1G220260. The TRITD1Av1G215640 and TRITD1Bv1G207110 were identified as *TduASN3.1*, and TRITD1Av1G222870 and TRITD1Bv1G220260 were identified as *TduASN3.2*. One *ASN* gene was identified on chromosome 3, TRITD3Av1G021350 on chromosome 3A; however, an unassigned gene, TRITD0Uv1G024260, was identified as the chromosome 3B homeologue as it showed 91.0 % identity to TRITD3Av1g021350 (Figure 3.6). All of the genes identified, except for TRITD0Uv1G024260, the B genome *TduASN2*, were already annotated as encoding asparagine synthetases in the Ensembl database.

The *TduASN1* A and B homeologues shared 97.4 % identity with each other, and both shared 99.8 % identity to their respective counterparts in hexaploid wheat (Figure 3.6). The B genome version of *TduASN1* showed 99.2 % identity to the D genome version of *TaASN1*, with the A genome only

showing 97.7 %. The *TduASN1* genes showed 83.9 – 84.1 % identity to the *TaASN2* genes. The *TduASN2* genes shared 91.4 % identity. The A genome *TduASN2* showed 94.7 % identity to the A genome *TaASN2*, with the unassigned gene, presumed to be the B genome homeologue, actually showing more similarity to the A genome *TaASN2* at 95.6 % identity. There was no *TaASN2* B genome homeologue to compare it to. The A and presumed B genome *TduASN2* genes showed 92.0 and 96.1 % identity to the D genome *TaASN2*.

The A and B genome *TduASN4* genes shared 92.7 % identity, showed 96.3 and 99.8 % identity to their *TaASN4* counterparts, and showed 93.3 and 96.9 % identity to the D genome *TaASN4*, respectively. Two genes were identified on both chromosomes 1A and 1B. The 1A genes shared 90.7 % identity, and the 1B genes shared 92.5 % identity. TRITD1Av1G215640, on chromosome 1A, showed greater similarity to the *TaASN3.1* genes than the *TaASN3.2* genes (94.3 - 98.2 % compared with 90.5 - 91.4 %) and TRITD1Av1G222870, the second chromosome 1A gene, showed the highest similarity to the *TaASN3.2* genes (93.2 - 97.1 % compared with 92.3 - 92.9 % to the *TaASN3.1* genes). Out of the chromosome 1B genes, TRITD1Bv1G207110 showed greater similarity to the *TaASN3.1* genes (94.5 - 97.7 %) than to the *TaASN3.2* genes (92.5 - 94.2 %), and TRITD1Bv1G220260 showed the reverse, with the greatest similarity shown to the *TaASN3.2* genes (95.7 - 98.8 % compared with 92.3 - 92.9 %). Consequently, TRITD1Av1G215640 and TRITD1Bv1G207110 were named as the *TduASN3.1* genes and TRITD1Av1G222870 and TRITD1Bv1G220260 annotated as *TduASN3.2*. The *Triticum durum* genes were also compared to the *Triticum dicoccoides*. The orthologues all shared over 96 % identity, which is expected given their close evolutionary history.

Figure 3.6. Similarity matrix of the *TaASN* and *TduASN* genes calculated from a nucleotide alignment of the genes generated using Geneious. The values presented are the percentage of bases that are identical, expressed to one decimal place.

3.2.3.4. The Triticea: Goat grass (*Aegilops tauschii*)

Aegilops tauschii, or Tausch's goat grass or rough-spike hard grass, is a diploid wheat species (DD) that is accepted to be the progenitor of the D-genome of bread wheat (McFadden and Sears, 1946). The reference is based on the subspecies *Aegilops tauschii* subsp. *strangulata* (Aet v4.0, INSDC Assembly; Luo *et al*., 2017b).

Table 3.4. Gene names (Raffan and Halford, 2021), Ensembl gene identifiers and wheat orthologues of the five *ASN* genes identified in *Aegilops tauschii*.

Gene name	Gene ID	Wheat orthologue
AetASN1	AET5Gv20393100	*TaASN1*
AetASN12	AET3Gv20170100	*TaASN2*
AetASN3.1	AET1Gv20919800	*TaASN3.1*
AetASN3.2	AET1Gv21000200	*TaASN3.2*
AetASN4	Aet4Gv20505300	*TaASN4*

Five *ASN* genes were identified in *Aegilops tauschii*, with single genes on chromosomes 3 (AET3Gv20170100), 4 (Aet4Gv20505300) and 5 (AET5Gv20393100), and two genes on chromosome 1 (AET1Gv21000200 and AET1Gv20919800), matching the gene family organisation seen in the other wheat species (Table 3.4). The two genes on chromosome 1, AET1Gv21000200 and AET1Gv20919800, shared 92.7 % identity, and the genes on chromosomes 5, AET5Gv20393100, and 3, AET3Gv20170100, shared 84.5 % identity (Figure 3.7). The genes showed high levels of similarity with the corresponding hexaploid wheat genes, with AET5Gv20393100, the gene on chromosome 5, and AET3Gv20170100, on chromosome 3, sharing 100 % identity with the D genome homeologues of *TaASN1* and *TaASN2*, respectively. Aet4Gv20505300, on chromosome 4, showed 99.9 % identity to the D genome homeologue of *TaASN4*. Out of two genes on chromosome 1, AET1Gv21000200 and AET1Gv20919800, the latter showed greater similarity to *TaASN3.1* (99.6 % for the D genome *TaASN3.1* compared with 92.9 % for the D genome *TaASN3.2*), while the former showed greater similarity to *TaASN3.2* (97.4 % for the D genome *TaASN3.2* gene compared with 92.7 % for the D genome *TaASN3.1*).

Figure 3.7. Similarity matrix of the *TaASN* and *AetASN* genes calculated from a nucleotide alignment of the genes generated using Geneious. The values presented are the percentage of bases that are identical, expressed to one decimal place.

3.2.3.5. The Triticea: Einkorn wheat (*Triticum urartu*)

Triticum urartu is a diploid wheat species (AA) which is commonly believed to be the progenitor of the A genome of both bread wheat and pasta wheat (Dvořák *et al*., 1993). The reference sequence is based on the accession G1812; however, no chromosomes were assembled for this genome, only scaffolds (ASM34745v1, INSDC Assembly; Ling *et al*., 2013). This meant that it was not possible to ascertain the chromosomal positioning of the identified genes.

Eight genes were initially identified in the study: TRIUR3_05036, TRIUR3_11865, TRIUR3_15772, TRIUR3_21196, TRIUR3_27580, TRIUR3_27838, TRIUR3_33442 and TRIUR3_34781 (Table 3.5; Figure 3.8). The comparison of the eight genes showed that TRIUR3_11865 and TRIUR3_15772 were the most similar (80.3 %), while TRIUR3_27580 and TRIUR3_05036 also showed high similarity (73.9 %). Three genes, TRIUR3_27838, TRIUR3_33442 and TRIUR3_34781, were much more divergent than any of the other *ASN*s in the analysis. When investigated, the genes did not show the expected intro-exon pattern (Section 3.3.1); in fact, the genes contained no introns. A BLAST search from the NCBI database indicated the genes were of bacterial origin. The TRIUR3_34781 and TRIUR3_33442 were identical to bacterial *ASNB* genes (Genbank KYP89277.1, KMV69586.1 for TRIUR3_34781; KMV69586.1 for TRIUR3_33442) from a bacterial symbiont of *Frankliniella occidentalis* (western flower thrips), and TRIUR3_27838 showed high similarity to an *ASNB* gene from a S*phingobacterium* (Genbank WP_116774391.1). These genes are clearly not from *Triticum urartu* and are likely the result of contamination of the DNA, leaving five identifiable *ASN* genes (Figure 3.4).

Table 3.5. Gene names (Raffan and Halford, 2021), Ensembl gene identifiers and wheat orthologues of the five *ASN* genes identified in *Triticum urartu*.

Gene name	Gene ID	Wheat orthologue
TuASN1	TRIUR3_11865	*TaASN1*
TuASN12	TRIUR3_15772	*TaASN2*
TuASN3.1	TRIUR3_27580	*TaASN3.1*
TuASN3.2	TRIUR3_05036	*TaASN3.2*
TuASN4	TRIUR3_21196	*TaASN4*

TRIUR3_11865 was the most similar to *TaASN1* (93.3 - 95.6 %), while TRIUR3_15772 was most similar to *TaASN2* (90.9 – 93.4 %) and TRIUR3_27580 was the most similar to *TaASN3.1* (81.6 – 84.6 %). TRIUR3_05036 was the most similar to the *TaASN3.2* genes (90.2 - 91.8 %) and TRIUR3_21196 was most similar to *TaASN4* (88.4 - 89.9 %), which matches with the expected gene family pattern.

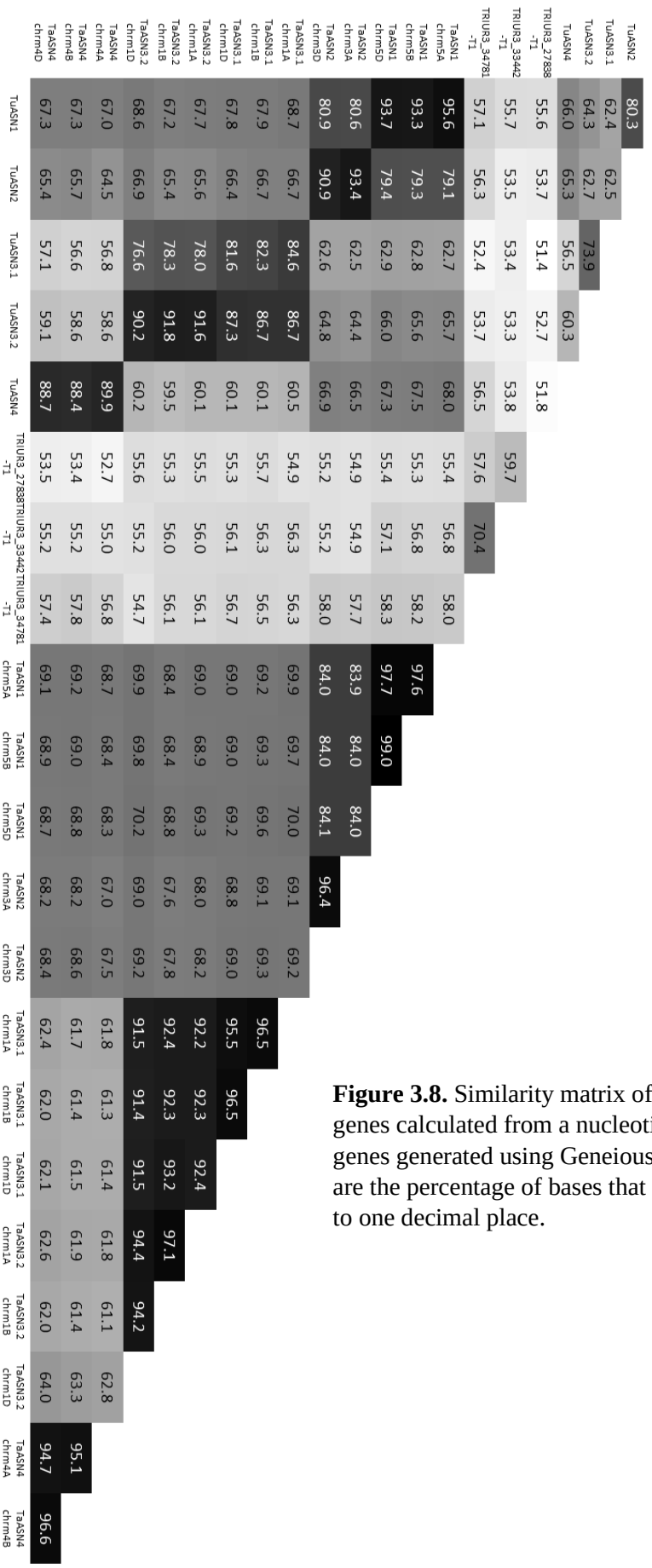


Figure 3.8. Similarity matrix of the *TaASN* and *TuASN* genes calculated from a nucleotide alignment of the genes generated using Geneious. The values presented are the percentage of bases that are identical, expressed to one decimal place.

3.2.3.6. The Triticea: Barley *(Hordeum vulgare)*

Barley (*Hordeum vulgare*) is a diploid species with, like its close relative, wheat, seven chromosomes in its haploid genome. The reference sequence is based on the variety Morex (IBSC v2, INSDC Assembly; International Barley Genome Sequencing Consortium (Mayer *et al*., 2012). Two *ASN* genes were initially described in barley (Møller *et al*., 2003), with five genes later being identified and annotated as *HvASN1* - *HvASN5* (Avila-Ospina *et al*., 2015) (Table 3.5).

Table 3.6. Gene names (Raffan and Halford, 2021), Ensembl gene identifiers, wheat orthologues, and the gene names given in Avila-Ospina *et al*., (2015) of the five *ASN* genes identified in barley *(Hordeum vulgare)*.

Gene name	Gene ID	Wheat orthologue	Gene name in Avila-Ospina *et al.*, (2015)
HvASN1	HORVU5Hr1G048100	*TaASN1*	*HvASN1*
HvASN12	HORVU3Hr1G013910	*TaASN2*	*HvASN2*
HvASN3.1	HORVU1Hr1G084370	*TaASN3.1*	*HvASN3*
HvASN3.2	HORVU1Hr1G092110	*TaASN3.2*	*HvASN4*
HvASN4	HORVU4Hr1G056240	*TaASN4*	*HvASN5*

Five *ASN* genes were identified in this analysis (Table 3.6), with one gene, HORVU3Hr1G013910 on chromosome 3, already annotated as an *ASN2* (UniProtKB Q84LA5; https://www.uniprot.org/uniprot/Q84LA5). The other genes were present on chromosome 5 (HORVU5Hr1G048100), chromosome 4 (HORVU4Hr1G056240), with two genes present on chromosome 1 (HORVU1Hr1G084370 and HORVU1Hr1G092110), matching the gene organisation seen in wheat. As expected, the chromosome 1 genes shared the highest identity (92.7 %), which was followed by the genes on chromosomes 5 and 3, which share 80.1 % (Figure 3.9). When compared to the *TaASN* genes, the *HvASN2* gene (HORVU3Hr1G013910) was very similar to the *TaASN2* genes, as was expected, with 91.8 and 92.3 % shared identity with the A and D genome homeologues. The chromosome 5 gene, HORVU5Hr1G048100, showed 96.3 - 96.5 % identity to the *TaASN1* genes, and the chromosome 4 gene, HORVU4Hr1G056240, showed 94.6 - 95.3 % identity to the *TaASN4* genes. Out of the genes on chromosome 1, HORVU1Hr1G084370 was more similar to the *TaASN3.1* genes than to the *TaASN3.2* genes (94.9 - 96.1 % compared with 92.0 - 92.4 %) and HORVU1Hr1G092110 was more similar to the *TaASN3.2* genes than the *TaASN3.1* genes (94.5 - 96.8 % compared with 92.3 - 92.8 %).

Figure 3.9. Similarity matrix of the *TaASN* and *HvASN* genes calculated from a nucleotide alignment of the genes generated in Geneious. The values presented are the percentage of bases that are identical, expressed to one decimal place.

3.2.3.7. The Triticea: Rye (*Secale cereale*)

Rye is a diploid species that diverged from wheat relatively recently, between three and seven million years ago (Marcussen *et al*., 2014). The genome data are based on the inbred line, Lo7 (Bauer *et al*., 2017) and was accessed at the GrainGenes website (https://wheat.pw.usda.gov/GG3/node/435). The sequence data are presented as scaffolds, with no assembly into chromosomes, so no positional information was available.

Table 3.7. Gene names (Raffan and Halford, 2021), the scaffold IDs from GrainGenes (Bauer *et al*., 2017) and wheat orthologues of the five *ASN* genes identified in rye (*Secale cereale*).

Gene name	Scaffold ID	Wheat orthologue
ScASN1	370516	*TaASN1*
ScASN2	174491	*TaASN2*
ScASN3.1	3445	*TaASN3.1*
ScASN3.2	1245, 81708, 36651	*TaASN3.2*
ScASN4	527072	*TaASN4*

Five asparagine synthetase genes were identified in the genomic data (Table 3.7). Four of these were each located on a single scaffold (3445, 174491, 370516 and 527072); however, one gene was located across three scaffolds (1245, 81708 and 36651).

The highest similarity was between the gene located on 3445 and the tri-scaffold gene (92.5 %). The genes on scaffold 370516 and 174491 shared 84.0 % identity (Figure 3.10). When compared to the *TaASN* genes, it was clear that the *ScASN* family followed the same pattern seen in the *TaASN* family. The scaffold 370516 gene is an *ASN1* (97.3 - 98.1 % identity to the *TaASN1* genes), the scaffold 174491 gene is an *ASN2* (95.1 % and 96.1 % identity to the 3A and 3D *TaASN2* genes), the scaffold 527072 gene is an *ASN4* gene (94.9 – 97.0 % identity to the *TaASN4* genes), and the scaffold 3445 and the tri-scaffold gene are the *ASN3* genes. The scaffold 3445 gene showed greater similarity to the *TaASN3.1* genes than the *TaASN3.2* genes (95.9 - 97.6 % compared to 92.4 - 92.9 %), while the tri-scaffold gene showed greater similarity to the *TaASN3.2* genes than the *TaASN3.1* genes (95.5 - 97.7 % compared to 92.5 - 93.1 %).

Figure 3.10. Similarity matrix of the *TaASN* and *ScASN* genes calculated from a nucleotide alignment of the genes generated using Geneious. The values presented are the percentage of bases that are identical, expressed to one decimal place.

3.2.3.8. Brachypodium (*Brachypodium distachyon*)

Brachypodium is not a cultivated crop species but is a model species for the cereals. It is particularly suitable as a model species as it is diploid, with a small genome of approximately 355 megabases and a short life cycle (Bennett & Leitch, 2005; Draper *et al*., 2001). It is a close relative of the Triticeae; however, it only possesses five chromosomes rather than seven. The reference sequence, produced by the Joint Genome Institute (International Brachypodium Initiative, 2010), is based on the Bd21 strain.

Table 3.8. Gene names (Raffan and Halford, 2021), Ensembl gene identifiers and wheat orthologues of the three *ASN* genes identified in *Brachypodium distachyon.*

Gene name	Gene ID	Wheat orthologue
BdASN1	BRADI_4g45010v3	*TaASN1*
BdASN3	BRADI_2g21050v3	*TaASN3.1/ TaASN3.2*
BdASN4	BRADI_1g65540v3	*TaASN4*

Genes were identified on chromosome 1 (*BRADI_1g65540v3*), chromosome 2 (*BRADI_2g21050v3*) and chromosome 4 (*BRADI_4g45010v3*) (Table 3.8). The genes did not show high levels of identity (Figure 3.11), with the greatest similarity seen between the chromosome 1 and chromosome 4 genes (70. 1 % identity). The chromosome 2 and chromosome 4 genes showed the lowest similarity, at 62.5 % identity.

When compared to the *TaASN* genes, the chromosome 1 *BdASN* gene, BRADI_1g65540v3, was the most similar to the *TaASN4* genes (86.9 – 87.8 %), the chromosome 2 *BdASN* gene, BRADI_2g21050v3, was most similar to the *TaASN3s* (89.6 - 90.2 % for *TaASN3.1*; 89.1 - 89.8 % for *TaASN3.2*), and the chromosome 4 gene, BRADI_4g45010v3, was most similar to the *TaASN1s* (90.7 – 91.0 %). BRADI_4g45010v3 also showed high levels of similarity to the two *TaASN2* genes, 84.4 and 84.7 % identity for the A and D homeologues respectively; however, there was no *TaASN2* orthologue in brachypodium.

Figure 3.11. Similarity matrix of the *TaASN* and *BdASN* genes calculated from a nucleotide alignment of the genes generated using Geneious. The values presented are the percentage of bases that are identical, expressed to one decimal place.

3.2.3.9. The Panicoideae: maize (*Zea mays*)

Maize (*Zea mays*) has a large genome (approximately 2.4 gigabases), with a haploid chromosome number of 10. The reference sequence is based on the cultivar B73 (B73 RefGen_v4, INSDC Assembly; Gramene; Jiao *et al*., 2017). Todd *et al*. (2008) identified four *ZmASN* genes, naming them *ZmASN1-4*. The analysis of the genome data confirmed these four *ZmASN* genes. Zm00001d045675 (annotated as *ZmASN1* by Todd *et al*. (2008)) and Zm00001d047736 (annotated as *ZmASN4*) were located on chromosome 9; Zm00001d028750 (annotated as *ZmASN3*) on chromosome 1; and Zm00001d044608 (annotated as *ZmASN2*) on chromosome 3 (Table 3.9). Three partial genes were also identified in the analysis. Two of these, Zm00001d031563 and Zm00001d028766, were located on chromosome 1, and the last one, Zm00001d010355, was located on chromosome 8.

Table 3.9. Gene names (Todd *et al*., 2008), Ensembl gene identifiers and wheat orthologues of the eight *ASN* genes identified in *Zea mays*.

Gene Name	Gene ID	Wheat orthologue
ZmASN2	Zm00001d044608	*TaASN1*
ZmASN1 (truncated)	Zm00001d010355	*TaASN1*
ZmASN1 (truncated)	Zm00001d031563	*TaASN1*
ZmASN1 (truncated)	Zm00001d028766	*TaASN1*
ZmASN1	Zm00001d045675	*TaASN3.1/TaASN3.2*
ZmASN3	Zm00001d028750	*TaASN4*
ZmASN4	Zm00001d047736	*TaASN4*

ZmASN3 and *ZmASN4* were the most similar out of the full-length genes, sharing 86.4 % identity, with low levels of similarity between the rest of the genes (55.4 – 69.8 %) (Figure 3.12). The highest similarity out of all the genes was between two of the partial genes, Zm00001d031563 and Zm00001d010355, which shared 91.6 % identity, with Zm00001d028766 sharing 79.0 and 83.5 % identity, respectively, to those two genes. All three truncated genes were most similar to *ZmASN2* (80.6 – 83.8 %), despite being annotated as *ZmASN1* homologues in the Ensembl database.

When compared to the *TaASN* genes, *ZmASN4* was the most similar to *TaASN4* (76.1 – 76.3 %); however, *ZmASN3* showed greater similarity to the *TaASN4* homeologues (85.0 - 85.5 %). It is likely the two genes arose from a single ancestral *ASN4* gene (section 4.3.2; Figure 3.12). The *ZmASN2* gene was most similar to *TaASN1* (77.7 – 78.1 %), followed by *TaASN2* (75.9 – 76.0 %), and *ZmASN1* was similar to the *TaASN3* genes (83.0 – 84.0 %).

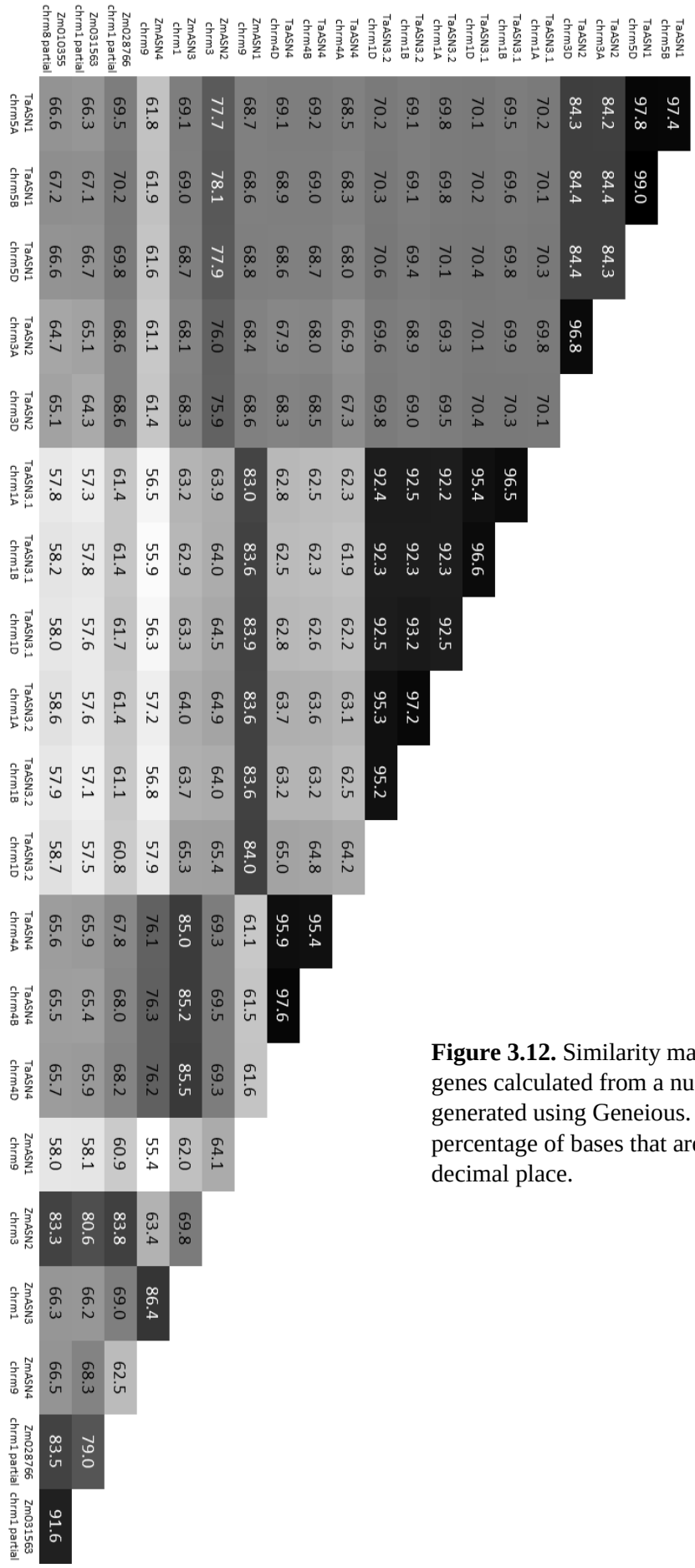


Figure 3.12. Similarity matrix of the *TaASN* and *ZmASN* genes calculated from a nucleotide alignment of the genes generated using Geneious. The values presented are the percentage of bases that are identical, expressed to one decimal place.

3.2.3.10. The Panicoideae: Sorghum (*Sorghum bicolor*)

Sorghum (*Sorghum bicolor*) is a close relative of maize. It is diploid, with a haploid chromosome number of 10. The reference sequence is based on the variety BTx623 (Sorghum_bicolor_NCBIv3, INSDC Assembly GCA_000003195.3; McCormick *et al*., 2018).

Table 3.10. Gene names (Raffan and Halford, 2021), Ensembl gene identifiers and wheat orthologues of the three *ASN* genes identified in *Sorghum bicolor*.

Gene name	Gene ID	Wheat orthologue
SbASN1	SORBI_3005G003200	*TaASN1*
SbASN3	SORBI_3010G110000	*TaASN3.1/ TaASN3.2*
SbASN4	SORBI_3001G406800	*TaASN4*

Three *ASN* genes were identified: SORBI_3001G406800 on chromosome 1, SORBI_3005G003200 on chromosome 5, and SORBI_3010G110000 on chromosome 10 (Table 3.10). SORBI_3001G406800 and SORBI_3005G003200 showed the most similarity (68.8 %), with SORBI_3010G110000 showing 62.0 – 63.4 % identity to the other two genes (Figure 3.10). SORBI_3010G110000 was most similar to the *TaASN3* genes, sharing 83.2 - 84.6 % nucleotide sequence identity, while SORBI_3001G406800 shared 83.1 – 84.1 % identity with the *TaASN4* homeologues. SORBI_3005G003200 was similar to *TaASN1,* followed by *TaASN2*, sharing 77.4 – 77.8 % and 75.7 – 75.8 % sequence identity, respectively, with those genes.

The *SbASN* genes were compared to the *ZmASNs* and, despite their different chromosomal locations, the genes between the species where very similar, reflecting their close evolutionary history (Figure 3.13). *SbASN1* shared 87.3 % identity with *ZmASN2* (75.5 – 77.0 % to the truncated *ZmASN* genes), *SbASN3* shared 95.2 % identity with *ZmASN1* and *SbASN4* shared 94.4 % identity with *ZmASN3*.

Figure 3.13. Similarity matrix of the *TaASN* and *SbASN* genes calculated from a nucleotide alignment of the genes generated using Geneious. The values presented are the percentage of bases that are identical, expressed to one decimal place.

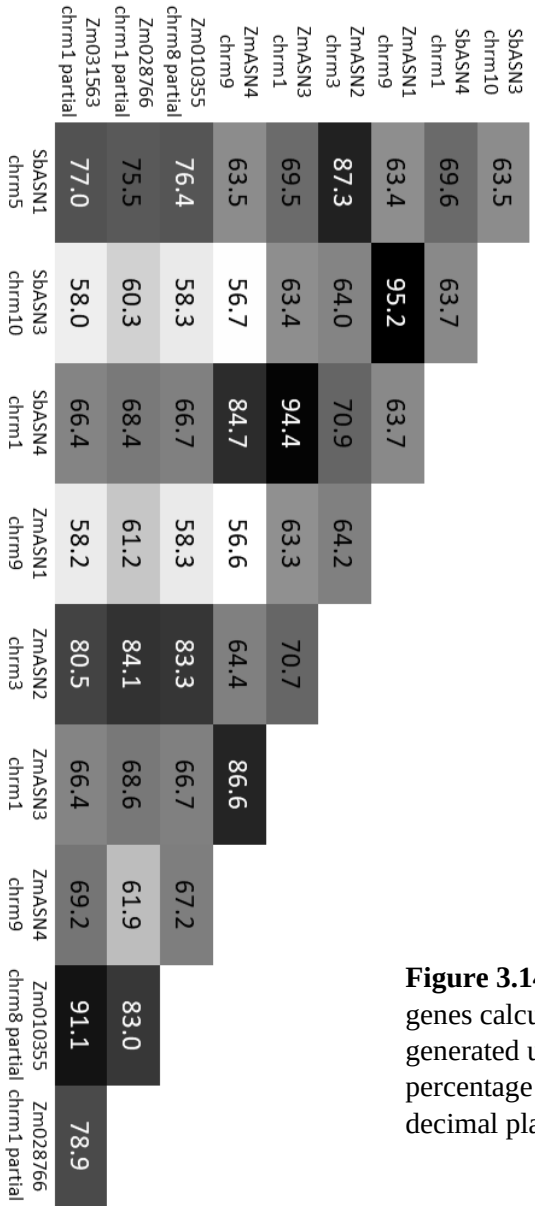


Figure 3.14. Similarity matrix of the *ZmASN* and *SbASN* genes calculated from a nucleotide alignment of the genes generated using Geneious. The values presented are the percentage of bases that are identical, expressed to one decimal place.

3.2.3.11. Rice (*Oryza sativa)*

There are two commonly cultivated rice subspecies, *indica* and *japonica* (Kawahara *et al*., 2013; Yu *et al*., 2002), both of which are diploid. Rice is a member of the Ehrhartoideae (also known as Oryzoideae) subfamily, which diverged from the Panicoideae and Pooideae line between 40 and 54 million years ago. Both the *indica* and *japonica* subspecies have been investigated here.

Table 3.11. Gene names (Raffan and Halford, 2021), Ensembl gene identifiers and wheat orthologues of the two *ASN* genes identified in *Oryza sativa* subsp. *indica* and *japonica.*

Existing annotation	Gene ID in Ensembl	Rice subspecies	Wheat orthologue
OsASN1	BGIOSGA010942	*indica*	*TaASN4*
OsASN1	Os03g0291500	*japonica*	*TaASN4*
OsASN2	BGIOSGA021489	*indica*	*TaASN3.1/TaASN3.2*
OsASN2	Os06g0265000	*japonica*	*TaASN3.1/TaASN3.2*

Two *ASN* genes in rice had been identified previously. The first was identified by Watanabe *et al*. (1996) then further characterised by Nakano *et al*. (2000) and is now known as *OsASN2*. The second was identified through a BLASTN search of the first draft rice genome (Møller *et al*., 2003) and confirmed by Sakai *et al*. (2013) and Ohashi *et al*. (2015), who annotated it as *OsASN1*.

Only two genes were identified in this analysis (Table 3.11). BGIOSGA010942 *(indica)* and Os03g0291500 *(japonica)* were identified on chromosome 3 and have previously been annotated as *OsASN1*. The other gene (BGIOSGA021489 and Os06g0265000 for *indica* and *japonica*, respectively) was located on chromosome 6 and previously identified as *OsASN2*. BGIOSGA021489 and Os06g0265000 share 100 % identity and encode identical 591 amino acid proteins; however, a difference is seen between the *OsASN1* genes (97.4 % identity) (Figure 3.15). BGIOSGA010942 was shown to be missing 42 bp at the 5' end of the coding sequence due to a different translation start sites. Thus, BGIOSGA010942 encodes a 590 amino acid protein and Os03g0291500 encodes a 604 amino acid protein. When comparing the *OsASN* genes, only 63.3 % nucleotide sequence identity was seen between Os03g0291500 (*OsASN1*) and the *OsASN2* genes, and 62.0 % identity was seen between Os06g0265000 and the *OsASN2* genes.

Comparing the *OsASN* genes with the *TaASN* genes, the *OsASN1* genes showed the highest similarity to the *TaASN4* genes (83.5 – 86.3 %), and the *OsASN2* genes showed the highest similarity to the *TaASN3* genes (86.3 – 87.2 %). This means that rice does not possess a group 1 or a group 2 *ASN* gene.

Figure 3.15. Similarity matrix of the *TaASN* and *OsASN* genes calculated from a nucleotide alignment of the genes generated using Geneious. The values presented are the percentage of bases that are identical, expressed to one decimal place.

3.2.4. Assignment of all the cereal asparagine synthetase genes to groups

This analysis confirmed the four *ASN* groups, with group 3 subdivided (3.1 and 3.2). Genes were allocated into the different groups based on their similarity to the *TaASN* genes and numbered to match the wheat annotations (if not previously annotated), and this assignment was confirmed through a phylogenetic tree of all the genes (Figure 3.16). The analysis also showed that the genes which had been assigned to group1 based on their similarity matrices in maize and sorghum (Figures 3.12 and 3.13) actually clustered with the group 4 genes (Figure 3.16), showing a discrepancy between the tree topology and the group assignments. The genes were assigned to group 1 based on their similarity matrices (Figures 3.12 and 3.13), with the sorghum gene (*SbASN1*) showing 77.4 – 77.8 % similarity to *TaASN1*, but only 68.8 – 68.9 % similarity to *TaASN4*. Similarly, the maize (*ZmASN2*) showed 77.7 – 78.1 % similarity to TaASN1, and 69.3 – 69.5 %. However, the genes did cluster to the Group 4 genes, with a fairly high level of confidence. This suggests that the maise and sorghum group 1 genes may have diverged from the group 4 genes much less than the brachypodium and Triticea group 1 and 2 genes.

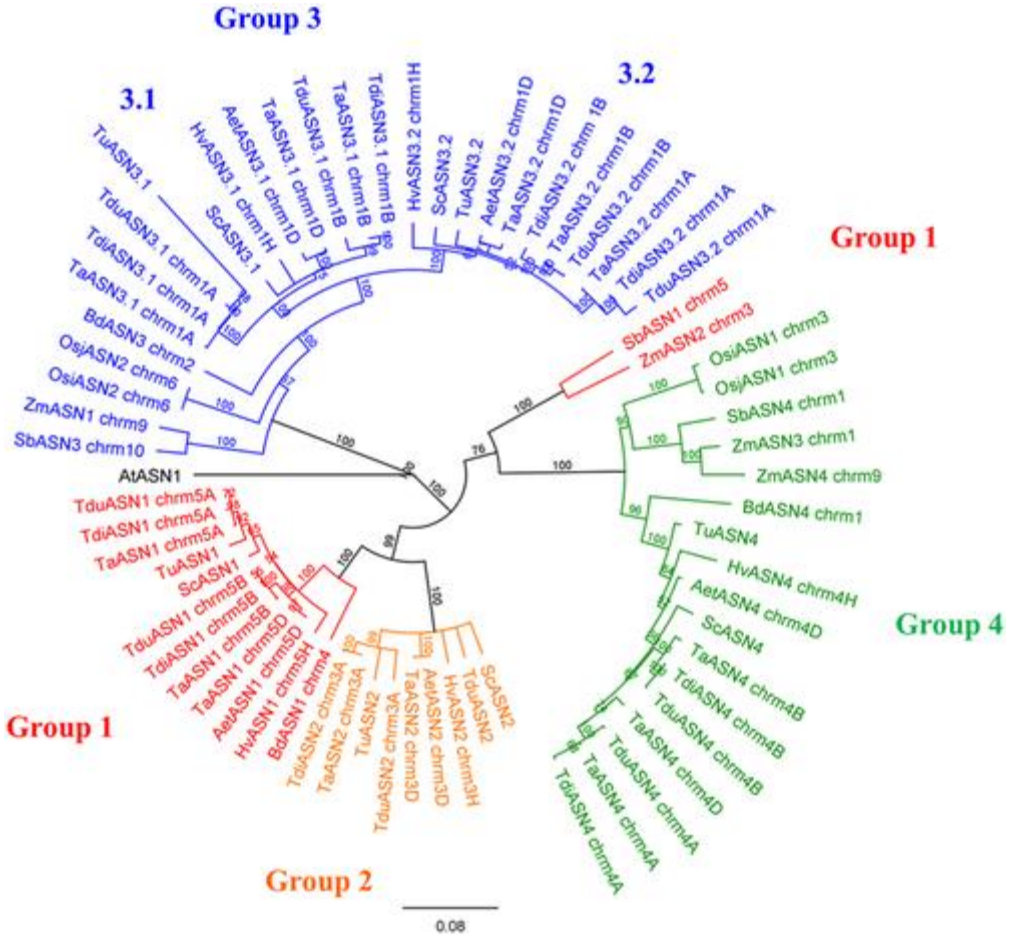


Figure 3.16. Phylogenetic tree of all the asparagine synthetase genes analysed in this study, shown with prefixes: Ta (bread wheat; *Triticum aestivum*), Tdi (emmer wheat; *Triticum dicoccoides*), Tdu (pasta wheat; *Triticum durum*), Tu (einkorn wheat; *Triticum urartu*), Aet (Tausch's goatgrass; *Aegilops tauschii*), Hv (barley; *Hordeum vulgare*), Sc (rye; *Secale cereale*), Zm

(maize; *Zea mays*), Sb (sorghum; *Sorghum bicolor*), Bd (brachypodium; *Brachypodium distachyon*) and Os (rice; *Oryzae sativa*). The branch labels show the consensus support in the clade from bootstrapping, shown as a percentage. The scale bar represents the length of the branches, expressed in units of substitutions per site of the sequence alignment. The Arabidopsis (*Arabidopsis thaliania*) gene, *AtASN1*, was used as an outgroup to anchor the tree.

3.2.4.1. Evolutionary development of the *ASN* gene families in cereal species

The genomic analysis confirmed that the wheat asparagine synthetase gene family structure, comprising five genes in four groups, is only found in the Triticeae. Although maize has a second group 4 gene, and some truncated genes, it, along with brachypodium and sorghum, does not possess a group 2 gene and has only one group 3 gene. Rice does not possess either a group 1 or a group 2 gene. Group 2 genes are unique to the Triticeae, which is important given that the group 2 genes are expressed seed-specifically. The evolutionary development of the *ASN* gene family in cereals seems to have begun with an initial duplication event of a single ancestral gene and the development of the group 3 and group 4 genes (Figure 3.17). As rice only contains group 3 and group 4 genes, it is likely that the Ehrhartoideae diverged from the ancestral line of the Panicoideae and Pooideae at this point, and their gene family did not radiate further. The Group 4 gene was then duplicated to give rise to the Group 1 gene, leading to families comprising Group 1, 3 and 4 genes, as is seen in maize, sorghum and brachypodium.

The genes in the ancestral Triticeae line must have undergone a further two duplications to get to the five gene structure seen in modern wheat. The first of these is the duplication of the *ASN1* gene, leading to *ASN1* and *ASN2*, and then the duplication of the *ASN3* gene, resulting in *ASN3.1* and *ASN3.2*. This five gene structure was seen in all the Triticeae investigated, although some varieties of both bread wheat (*Triticum aestivum*) and emmer wheat (*Triticum dicoccoides*) were shown to lack a B genome homeologue of *TaASN2*.

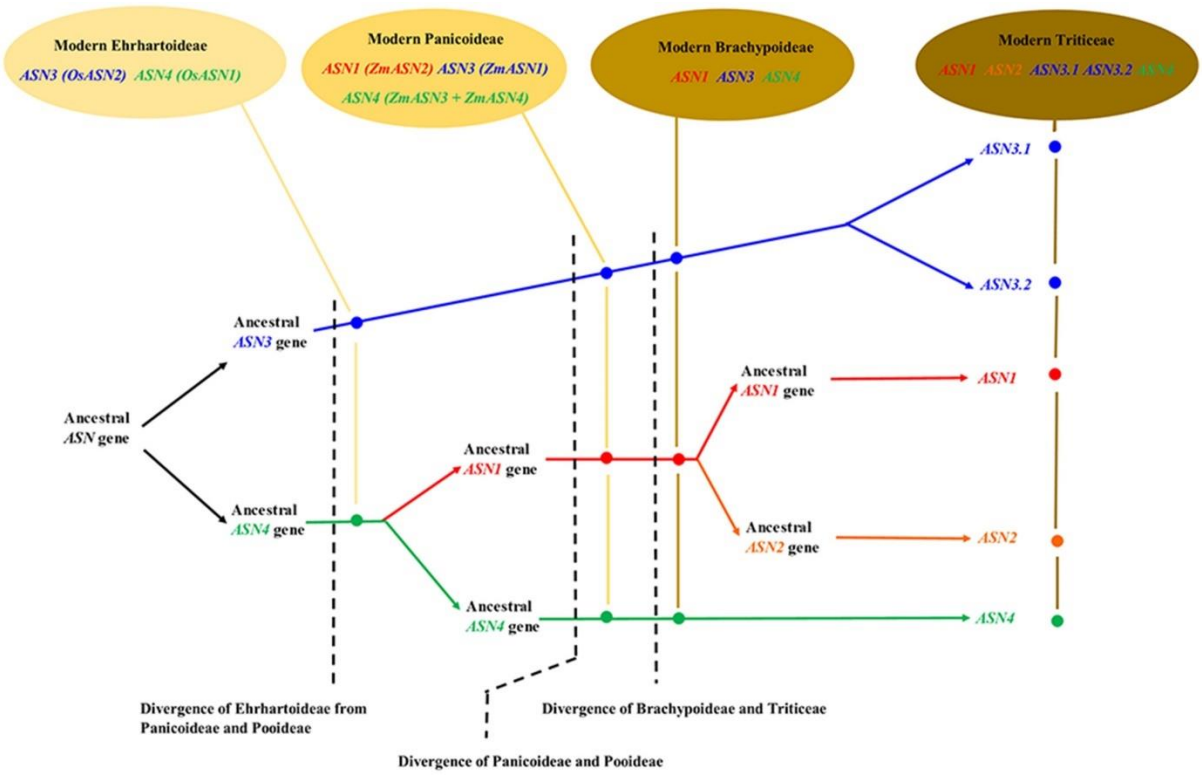


Figure 3.17. Diagram representing the evolution of asparagine synthetase genes in the Triticeae, Brachypodiae, Panicoideae and Ehrhartoideae (Raffan and Halford, 2021).

The Arabidopsis gene *AtASN1* (Lam *et al*., 1994) was used in the tree construction to anchor the genes. There are three *ASN* genes present in Arabidopsis (Lam *et al*., 1998), and it has been proposed (Gaufichon *et al*., 2010) that these genes fit into two different classes (I and II), with *AtASN1* in class I and the others in class II. A third class was proposed when asparagine synthetase genes were investigated in both monocotyledonous and dicotyledonous species (Duff, 2015), with a cluster of cereal genes emerging as a separate clade. This clade corresponds to the group 3 genes described in this chapter. Rice was shown to have class II and III genes, with class II corresponding to the group 4 genes described in this chapter, and class I corresponding to group 1. In contrast to the evolutionary path lain out in Figure 3.17, Duff (2015) proposed that the three classes of *ASN* genes developed before monocot and dicot plants divided. However, rice does not have a group 1 *ASN* gene, which would mean that rice would have once possessed an *OsASN1* gene but lost it if this model were correct. The genomic analyses presented in this chapter show that the cereal gene family has evolved considerably since the divergence of monocots and dicots, making the model shown in Figure 3.17 a better fit for the cereal data now available than one that tries to fit the cereal genes within a broader classification that attempts to include all of the other plant genes as well.

3.2.5. Related protein

The genes TraesCS5A02G081800, TraesCS5B02G084600 and TraesCS5D02G090800 (on chromosome 5A, 5B and 5D respectively), which share over 96 % cDNA identity, are annotated as asparagine synthetase genes in the Ensembl database. The genes are described in the database as encoding proteins containing glutamine amidotransferase 2 (IPR017932) (Mitchell *et al*., 2019) and asparagine synthetase (IPR001962) (Mitchell *et al*., 2019) domains, suggesting that the proteins are likely to possess asparagine synthetase activity; however, no cDNA alignment could be produced when attempting to align these genes to the previously identified *TaASN* genes using the Geneious alignment algorithm (Kearse *et al*., 2012). Furthermore, no Geneious alignment could be produced between these putative *ASN* genes and the *AtASN* genes from Arabidopsis.

The predicted protein sequence for the A genome homeologue was run through NCBI's Conserved Domain Database (CDD) (Marchler-Bauer *et al*., 2015) to assess the validity of the domain labelling. Two domain hits were returned. The first domain, the AsnB super family (Accession cl33852; E-value $4.66e^{-31}$) domain, spanned the entire 645 residues. The second domain, Gn_AT_II_novel (Accession cd03766; e-value $5.10e^{-54}$), was labelled as an asparagine synthetase-related domain and spanned the first 267 residues of the sequence. The domain is present in eukaryotes, but the function of this domain appears largely unknown. Residues 2 (cysteine), 99 (threonine; leucine/isoleucine/phenylalanine in matched sequences), 123-125 (asparagine, glycine, glutamate) and 138 (aspartate) were highlighted as important. The homeologues are expressed in wheat, albeit to a very low level (<10 transcripts per million), with the highest expression seen from TraesCS5B02G084600 in the developing embryo (Borrill *et al*., 2016; Ramírez-González *et al*., 2018). When the glutamine amidotransferase type 2 domain and the asparagine synthetase domain were extracted from TraesCS5A02G081800 and aligned to the *TaASN1* (5A) sequences, and the amino acids they encode, no alignment could be made, and no similarity was seen.

A BLAST search was performed on the NCBI database to ascertain what gene types this putative *ASN* gene was similar to. The first hit, with 96.95 % identity was to an asparagine synthetase domain-containing protein gene from *Aegilops tauschii*. This was followed by versions from *Sorghum bicolor*, *Zea mays* and *Oryza sativa*. No asparagine synthetase genes were identified, either plant or bacterial in origin. Corresponding genes were found in all the species investigated in this study, apart from *Oryza sativa* subsp. *japonica*, although a gene is present in the subspecies *indica* (Table 3.8). Despite this gene's presence in every plant species investigated, no bacterial or fungal counterparts could be established, reflecting the fact that the functional domain seemed confined to eukaryotes (Marchler-Bauer *et al*., 2015). It is concluded that this gene is not a true *ASN* gene.

Table 3.12. Genes encoding a protein previously identified as an asparagine synthetase-related domain identified in the species in this analysis.

Species	Gene ID
Triticum aestivum	TraesCS5A02G081800
	TraesCS5B02G084600
	TraesCS5D02G090800
Arabidopsis thaliana	AT2G03667
Triticum diccoides	TRIDC5AG011850
	TRIDC5BG013920
Triticum urartu	TRIUR3_31295
Aegilops tauschii	AET5Gv20219000
Hordeum vulgare	HORVU5Hr1G020510
Brachypodium distachyon	BRADI_4g03827v3
Sorghum bicolor	SORBI_3008G139800
Zea mays	Zm00001d030760
Oryza sativa subsp. *indica*	BGIOSGA035961

3.3. Discussion

At the start of this project, wheat was known to contain four classes of asparagine synthetase genes, *TaASN1–4* (Gao *et al*., 2016). This analysis established that there are single copies of *TaASN1*, *TaASN2*, and *TaASN4*, and two of *TaASN3*, and identified their chromosomal locations. The relatively simple structure of the gene family means that genetic interventions to reduce free asparagine accumulation and thereby acrylamide-forming potential in wheat grain are more likely to be successful.

The B genome homeologue of *TaASN2* is present in some hexaploid bread wheat varieties (including Cadenza, section 3.3.1), and it is also present in the genome of *Triticum durum* cv. Svevo, despite not being present in the reference sequence of their progenitor, *Triticum dicoccoides*. It is possible that some *Triticum dicoccoides* genotypes possess a B genome *ASN2* and that the hybridisation event with *Aegilops tauschii* that produced modern bread wheat occurred multiple times, with and without the *ASN2* B genome homeologue. It is also possible but less likely that when the hybridisation event occurred, *Triticum dicoccoides* possessed a B genome *TaASN2*, but that this gene was subsequently lost in some emmer and bread wheat genotypes. The B genome homeologue could also have been regained from an introgression from a related species, such as a wild wheat or rye.

This study highlighted the inconsistency in the pre-existing *ASN* gene family nomenclature across plant species. The five *TaASN* genes have been annotated *TaASN1*, *TaASN2*, *TaASN3.1*, *TaASN3.2* and *TaASN4* (Xu *et al*., 2018). Barley also contains five *HvASN* genes, orthologous to the *TaASN* genes, however these have been named *HvASN1 – HvASN5* (Avila-Ospina *et al*., 2015). As the gene family structure is the same in the two species, the nomenclature can be easily reconciled if *HvASN4* is rebranded as *HvASN3.2*, and *HvASN5* becomes *HvASN4*. The genes in maize, rice and wheat were

named in the order of their identification. However, the maize *ZmASN1* gene is a Group 3 gene, and the rice *OsASN1* gene is a group 4 gene. *ZmASN2* is a group 1 gene, with both *ZmASN3* and *ZmASN4* being group 4 genes, and *OsASN2* is a group 3 gene.

3.3.1. Expression of *ASN* genes

The implications of the differences in the asparagine synthetase gene family structures for free asparagine concentrations in the grain of the different species are currently unknown, and will depend in part on the expression patterns of the different genes. *ASN* gene expression has been shown to be highly variable in *Aegilops tauschii* seedlings (Nishijima, *et al*., 2016), dependant on genotype, with *AetASN1* generally being the most highly expressed. No *AetASN2* expression was detected in that study, suggesting that *AetASN2* could be grain-specific in its expression, like *TaASN2* (Gao *et al*., 2016), but expression in the grain has not yet been investigated. Two barley genes, *HvASN1* and *HvASN2*, the orthologues of *TaASN1* and *TaASN2*, also have different expression patterns (Møller *et al*., 2003). No *HvASN2* expression could be detected in the roots; however, it was detected in the leaves, which would make it different to *TaASN2*, with *HvASN1* expressed in both tissues. *HvASN3* and *HvASN4* were shown to be the most highly expressed genes in germinating barley (Zhang *et al*., 2016a) and barley seedlings (Mayer *et al*., 2012), with *HvASN2* showing very low levels of expression. The expression of the *HvASNs* has not been investigated in the grain.

During early development of the grain of brachypodium, *BdASN3* (*TaASN3* orthologue) has been shown to be the most highly expressed *ASN* gene (Davidson *et al*., 2012); however, by mid-development, *BdASN1* (*TaASN1* orthologue) was the dominant *ASN*, with expression higher in the embryo than in the endosperm. Broadly, this matches what is seen in bread wheat (Section 3.3.2), where *ASN* expression is higher in the embryo, although there is no *TaASN2* orthologue in brachypodium.

In the seeds of sorghum, *SbASN3* is the most highly expressed overall, but *SbASN1* shows higher expression in the embryo. Indeed, *SbASN1* shows embryo-specific expression (Davidson *et al*., 2012; Makita *et al*., 2015), which is even more narrow than the expression range of *TaASN2*. *SbASN3* and *SbASN4* show expression throughout the plant (Davidson *et al*., 2012). In maize, Todd *et al*. (2008) showed that *ZmASN1* (a group 3 gene) and *ZmASN4* (a group 4 gene) are expressed in all tissues, with *ZmASN3* (a group 4 gene) expression restricted to below-ground tissues and *ZmASN2* (group 1 gene) expression detectable in root, ear node, cob and seed tissues. The most highly expressed *ZmASN* gene in the grain is *ZmASN1* (Chen *et al*., 2017; Zhan *et al*., 2015); however, it shows high expression throughout the plant (Baute *et al*., 2015; Chang *et al*., 2012; Pang *et al*., 2019). *ZmASN2* was also expressed in the leaves, seedlings and 5-day-old caryopses (Baute *et al*., 2015; Chang *et al*., 2012; Pang *et al*., 2019) but to a much lower level, and *ZmASN3* was shown to be highly expressed in 5-day-old caryopses (Pang *et al*., 2019).

Rice only contains two *OsASN* genes, with *OsASN2* (a group 3 gene) most highly expressed in the seeds, as well as the shoots and root tips of developing seedlings (Davidson *et al*., 2012; Reynoso *et al*., 2018; Sakai *et al*., 2011; Zhang *et al*., 2014b). *OsASN1* (group 4) is expressed more in the anthers and carpels (Zhang *et al*., 2014b).

3.3.2. RNA-Seq

This analysis provided data on the genetic control of free asparagine accumulation in wheat grain and its response to sulphur supply. It confirmed that *TaASN2* was the most highly expressed *ASN* gene in the grain. The homeologues of many genes showed differential expression patterns, and this is particularly true for *TaASN2*, where the A genome homeologue was expressed to much higher degree than the D genome homeologue. No B genome homeologue could be identified in the study, indicating that the gene was either missing, or was not expressed; however, it is known that some wheat varieties are missing a B genome version of *TaASN2* (Xu *et al*., 2018).

The analysis confirmed that *TaASN2* was the logical target for genetic interventions, and interventions aimed at the A genome homeologue may be enough to generate a step change in grain free asparagine concentrations. The study also highlighted further targets, such as metabolic enzymes and signalling factors, for reducing the acrylamide-forming potential of wheat.

3.3.3. Targeted asparagine reduction in other species

This analysis showed that genetic interventions and technology employed in wheat would not necessarily transfer across to non-Triticeae cereals, without extensive prior investigation and characterisation of the *ASN* gene family. Although clearly related, the structure of the *ASN* families varies across the cereal species investigated, with the number of full-length *ASN* genes ranging from 2 - 5, and the representation of the different *ASN* groups also different. All species investigated possessed at least one Group 3 and Group 4 gene; however, the Group 1/2 genes were absent in rice and the gene targeted for genetic intervention in wheat, *TaASN2*, was found to be unique to the Triticea. Further investigation is needed to assess whether the Group 2 genes are grain-specific in the other Triticea members, as is seen in hexaploid bread wheat. Previous genetic interventions to knock-out *ASN* genes in rice and potato were met with undesirable phenotypes (Luo *et al*., 2019; Rommens *et al*., 2008); however, the unique presence of the *TaASN2* gene, and its grain-specific expression (Gao *et al*., 2016), presents a singular opportunity for reducing the acrylamide-forming potential of wheat grain.

Chapter 4

gRNA Generation and Protoplast Trial System

4.1. Introduction

Wheat transformation is a time-consuming process. Thus, designing several guide RNAs (gRNAs) and testing them first in a protoplast system is a sensible approach. This work aimed to design gRNAs to introduce a targeted mutation in the *TaASN2* gene in wheat (*Triticum aestivum*) and test them in a cv. Cadenza protoplast system. Guide RNAs that worked in the protoplast system would be selected for the generation of edited wheat plants. The protoplast trial assay was not successful; however, a multiplexed gRNA construct, containing four gRNAs, was designed to maximise the chances of successful editing.

4.1.1. gRNA design

In genome editing using the CRISPR/Cas9 system, the Cas9 nuclease is directed to a target site by an engineered, sequence-specific, single guide RNA (gRNA, sometimes referred to as sgRNA). Thus, the synthetic gRNA is one of the main components of the CRISPR/Cas9 genome editing system. It comprises two main components: a scaffold sequence for Cas9 binding, and a user-defined target sequence corresponding to a desired target region of the DNA. The target region must be directly upstream of a protospacer adjacent motif (PAM). The conventional (or classic) gRNA structure (Hsu *et al*., 2013; Anders *et al*., 2014; Nishimasu *et al*., 2014; Jinek *et al*., 2012) is a modified version of the native bacterial CRISPR RNA (crRNA)–transactivating crRNA (tracrRNA) duplex.

The site-specific targeting of the CRISPR/Cas9 system is determined by the first 20 nucleotides of the gRNA. This distinguishes the CRISPR/Cas9 method from techniques in which the DNA recognition is provided by a protein binding domain, as with zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) (Jinek *et al*., 2012). The targeting specificity of the CRISPR/Cas9 system (Wu *et al*., 2014a) is affected by the specificity incorporated in the Cas9/gRNA complex, which can vary for different gRNA sequences, and the relative abundance of the effective Cas9/gRNA complexes in the cell.

Many factors contribute to this targeting specificity. The PAM is one factor, and it is required to be immediately next to the 3' end of the target sequence. The PAM sequence is recognized by a specific domain in the Cas9 protein (Nishimasu *et al*., 2014) and varies across bacterial species (Garneau *et al*., 2010; Zhang *et al*., 2013). For the widely-used Cas9 from *Streptococcus pyogenes*, which is the Cas9 used in this work, the PAM is typically NGG, with no nucleotide bias at the first position.

Mismatches are tolerated to a greater or lesser extent at different positions in the gRNA guide sequence (Fu *et al*., 2013; Pattanayak *et al*., 2013; Hsu *et al*., 2013). When tested *in vitro*, mismatches in the first 7 positions (furthest from the PAM motif) of the gRNA sequence were well tolerated and led to plasmid cleavage by the Cas9 nuclease; however, mismatches in the 10-12 base pairs in the

PAM-proximal region usually led to decreased or suppressed target cleavage activity (Jinek *et al*., 2012). Thus, this PAM-proximal 10-12 bases has been defined as the 'seed region for Cas9 cutting activity'(Cong *et al*., 2013; Jiang *et al*., 2013a), and although the definition of this seed region has been debated (Kuscu *et al*., 2014; Wu *et al*., 2014b; O'Geen *et al*., 2014; Pattanayak *et al*., 2013), particularly in genome-wide binding datasets, it is generally accepted that mismatches near the PAM region are less well tolerated (Wu *et al*., 2014a; Kuscu *et al*., 2014). This is particularly important when designing gRNAs to differentiate between member of gene families.

Changes in the gRNA sequence may also affect the concentration of the gRNAs in the cell (Wu *et al*., 2014b), such as by affecting the transcription or stability of the gRNA; they may also affect gRNA-Cas9 interactions. The chromatin structure at the target site can also affect accessibility for Cas9, PAM recognition and target binding, with a strong correlation between Cas9-bound sites and open chromatin; however, studies have suggested that chromatin accessibility is not a requirement for binding to the on-target site (Wu *et al*., 2014a; Wu *et al*., 2014b; Kuscu *et al*., 2014; Perez-Pinera *et al*., 2013), preferentially facilitating off-target binding, and so may not affect gRNA design. Additionally, there may be epigenetic effects on gRNA efficiency: specifically, more methylation is associated with less binding (Wu *et al*., 2014b); however, CpG methylation has been shown to have no effect on Cas9 cutting efficiency *in vitro* or *in vivo* (Hsu *et al*., 2013). Xie *et al*. (2015) observed that gRNA target sequences with higher GC contents had relatively high editing efficiencies, recommending selecting target sites with GC contents of about 50 - 70 %; however, it has been noted that this could lead to a higher risk of off-target effects (Tsai *et al*., 2015).

The gRNA target sequence length can also affect editing or targeting efficiency, and it has been shown that only 20 bp of the gRNA sequence are present in mature gRNAs *in vivo*, even if the length of the guide region is increased (Ran *et al*., 2013); so increasing the target guide sequence does not increase the Cas9/gRNA complex specificity. Truncating the gRNA guide sequence to 17 or 18 nucleotides, on the other hand, has been shown to dramatically increase target specificity (Fu *et al*., 2014), although the mechanism behind this is not well understood. It was suggested that the first 2 - 3 nucleotides, which are not required for target binding, may stabilise off-target binding.

4.1.2. Online tools for gRNA design

Being able to distinguish effective from ineffective gRNAs can greatly streamline an experiment and simplify interpretation of results, but in the case of wheat, maximising gRNA efficiency is particularly important because wheat transformation is costly and time-consuming. Many resources now exist to help design effective gRNAs (https://blog.addgene.org/how-to-design-your-grna-for-crispr-genome-editing; https://www.takarabio.com/learning-centers/gene-function/gene-editing/gene-editing-tools-and-information/how-to-design-sgrna-sequences). Several web-based tools are available that require a DNA input sequence, a genomic region or a gene to be targeted and an indication of the target

species. An algorithm specific to each tool outputs a list of candidate guide sequences with corresponding predicted off-target sites for each input (Wu *et al*., 2014a).

The tools vary in the methods they employ, but most aim to provide guide sequences that minimize the likelihood of off-target effects. Some tools incorporate data from previous systematic mutagenic studies or user-input penalties (for example, CASFINDER (Aach *et al*., 2014); ECRISPR (Heigwer *et al*., 2014)) to score off-targets individually, based on location and number of mismatches to the guide sequence to rank the potential for off-target effects.

Other tools have binary criteria for off-target effects, either to the entire guide region (Cas-OFFinder (Bae *et al*., 2014)) or to some defined region ((Ma *et al*., 2013); CasOT (Xiao *et al*., 2014); CHOPCHOP (Montague *et al*., 2014); FlyCRISPR (Gratz *et al*., 2014); sgRNAcas9 (Xie *et al*., 2014)), and potential guide sequences are generally ranked by a weighted sum of off-target scores, or by the number of off-targets.

Several tools consider factors beyond position and number of mismatches. Some tools (CASFINDER, for example) include the option to score off-targets with alternate PAMs based on the finding that Cas9 cleaves these sites with lower efficiency (Mali *et al*., 2013a; Hsu *et al*., 2013; Pattanayak *et al*., 2013; Ran *et al*., 2013). A tool can also consider the presence of SNPs and secondary structures in the potential guide sequence (Ma *et al*., 2013), which could impact targeting and the formation of the Cas9/gRNA complex (Makarova *et al*., 2011), the genomic context of the potential target sequence (e.g. exons, transcripts, CpG islands), which could impact the intended purpose of the gRNA (ECRISP (Heigwer *et al*., 2014); CHOPCHOP (Montague *et al*., 2014)), and the GC content, which could impact effectiveness of the gRNA (ECRISP (Heigwer *et al*., 2014); CHOPCHOP (Montague *et al*., 2014); sgRNAcas9 (Xie *et al*., 2014)).

Tools have also been designed for specific purposes, such as CRISPR-ERA (Liu *et al*., 2015a), which designs gRNAs for gene repression or activation, or for specific species or taxonomic groups, such as FlyCRISPR (Gratz *et al*., 2014). Most gRNA tools are solely focused on the *Staphylococcus aureus* Cas9 (Ran *et al*., 2015); however, as alternative nucleases become increasingly available, such as the nuclease Cpf1 (Zetsche *et al*., 2015), design tools that are compatible are starting to be developed, such as the Benchling design tool (https://www.benchling.com/crispr/).

Although many of these tools are designed for mammalian or animal research (Doench *et al*., 2016) (https://genhub.co/design; http://guides.sanjanalab.org/#/; https://eu.idtdna.com/site/order/designtool/index/CRISPR_SEQUENCE PrimerQuest® program, IDT, Coralville, Iowa, USA. Accessed 12 December, 2018), or only contain Arabidopsis sequences (https://www.atum.bio/eCommerce/cas9/input), tools do exist for the design of gRNAs in wheat or other cereal species (Montague *et al*., 2014; Bae *et al*., 2014; Haeussler *et al*., 2016; Heigwer *et al*.,

2014; Hough *et al*., 2016; https://horizondiscovery.com/en/products/tools/CRISPR-Design-Tool; https://design.synthego.com/#/).

Due to the variety and number of online tools available, and the differences between them, it is critical to use more than one tool during the gRNA design process and then choose gRNA sequences that consistently perform well. For this project, the RGEN program (Park *et al*., 2015) was chosen, using '*Hordeum vulgare* (Ensembl Plants 28)' as the reference genome because a wheat reference genome was not available when the project started, along with DESKGEN (Hough *et al*., 2016).

4.1.3. Multiplexing the gRNAs

The CRISPR/Cas9 system is limited by the expression and targeting efficiency of the gRNA, but also by the number of possible target sites for the CRISPR/Cas9 system, especially when working in polyploid species, such as wheat, or when targeting multigene families, where there might not be many sites available to target all the genes with a single gRNA. More importantly, multiple gRNAs designed with different target sequences can direct Cas9 to specific target sites (Cong *et al*., 2013; Wang *et al*., 2013), with Cas9 able to edit multiple loci simultaneously in the same individual. As plant transformation is time-consuming and expensive, particularly in the case of wheat, the use of multiple gRNAs simultaneously is preferable. This requires the construction and linking of several reliable expression cassettes in the same plasmid.

Xie *et al*. (2015) designed a robust CRISPR/Cas9 vector system for convenient and high-efficiency multiplexed genome editing in monocot and dicot plants, showing simultaneous targeting of multiple (up to eight) members of a gene family, multiple genes in a biosynthetic pathway, or multiple sites in a single gene. The system revolves around a single, synthetic, multiplexed gRNA polycistronic gene with a tandemly-arrayed tRNA–gRNA architecture. This system relies on the plant's endogenous tRNA-processing system, which precisely cleaves both ends of the tRNA precursor *in vivo*, generating multiple, fully-functional gRNAs. This means that multiple gRNAs can be generated from a single transgene. The system uses PCR-based procedures to rapidly generate multiple gRNA expression cassettes, which are assembled into the binary CRISPR/Cas9 vectors in one round of cloning by Golden Gate (GG) ligation (Engler *et al*., 2008).

4.1.4. Different gRNA scaffolds

In addition to alterations in the 5' end of the gRNA, modifications to the scaffold region of the gRNA can have an impact on gRNA efficiency, with changes in scaffold length leading to changes in gRNA expression levels (Hsu *et al*., 2013). To increase CRISPR/Cas9 efficiency, the *TaASN2*-targeting gRNAs were designed with an improved gRNA scaffold, which had been shown to have a significantly higher editing efficiency than those with the conventional scaffold (Dang *et al*., 2015).

The conventional scaffold structure (Hsu *et al*., 2013) has a shortened duplex and contains a continuous sequence of thymines, which is the pause signal for RNA polymerase III and thus could potentially reduce transcription efficiency (Dang at al., 2015). Changing these two components was initially shown to have little effect on knockout efficiency, although the duplex was extended by 10 bp, leading to the conclusion that the conventional structure was the most efficient architecture for genome editing purposes (Hsu *et al*., 2013). Despite this, it was then reported that the imaging efficiency of a dCas9 (a mutated version of Cas9 lacking nickase activity)–green fluorescent protein (GFP) fusion protein in cells was increased when a gRNA with a mutated, continuous sequence of Ts and extended duplex was used (Chen *et al*., 2013). Thus, it seemed likely that, if changing these two elements enhanced dCas9 binding to target sites, it might also increase the knockout efficiency of Cas9. Dang *et al*. (2015) investigated changing these two elements and generated an improved gRNA structure with improved knockout efficiency. This improved gRNA scaffold contains an extended duplex (lengthened by up to 5 bp, perhaps explaining the discrepancy with Hsu *et al*., 2013) and a mutation in the continuous sequence of thymines (the fourth thymine was changed to a cytosine or a guanine). The effects of the two key changes seemed different, with the mutation in the continuous sequence of thymines affecting gRNA production, explained as being a result of the affected pause signal, and the extended duplex leading to increased Cas9 functionality, either through increased binding or enhanced stability.

4.1.5. Off-target effects

As every position of the gRNA target sequence does not need to match the target site for effective editing, off-target effects may occur, where nonspecific and unintended editing occurs through sequence similarity to the target sequences. Given that the binding length is relatively short (20 bp, with 5-12 bp being critical) (Wu *et al*., 2014a), each gRNA potentially has multiple, maybe even thousands, of matches in the genome that are followed by the PAM motif, although this will vary between gRNA target sequences. Off target effects are not believed to be common in plants, with several studies reporting no off-target effects in a variety of species when sites with at least a conserved 12-nt seed sequence were investigated. These included Arabidopsis, tobacco (*Nicotiana benthamiana*), wheat, rice and sweet orange (*Citrus sinensis*) (Feng *et al*., 2014; Jia and Wang, 2014; Li *et al*., 2013; Nekrasov *et al*., 2013; Shan *et al*., 2013; Zhou *et al*., 2014). When investigating a single mismatch, a 1.6 % off-target mutation rate was reported in rice (Xie and Yang, 2013), which was five times lower than the on-target mutation rate.

For basic and applied research in plants, off-targeting may not be a critical problem. The risk of off-target edits occurring in plants by the CRISPR/Cas9 system may not be higher than that of the frequent somatic mutations that occur during tissue culture-based transformation, and is likely to be much lower than for other mutagenesis treatments, such as chemical or radiation mutagenesis. Furthermore, this risk of off-targeting can be minimized by selection of highly specific target

sequences (Lei *et al*., 2014; Xie *et al*., 2014), while unwanted off-target mutations, if they do occur, can be eliminated by crossing (and backcrossing) the mutant plants with their parental lines. As *TaASN2* is one of five *TaASN* genes in wheat (Chapter 3), there was a risk in this study of off-target effects occurring in other members of the *TaASN* gene family. This was particularly true in the case of *TaASN1*, which shares 92.1 – 93.0 % nucleotide sequence identity with *TaASN2* (Chapter 3) and the gRNAs in this project were designed with this in mind.

Target sequences with higher GC contents, whilst having a higher editing efficiency (Ma *et al*., 2015b), showed a higher potential for off-target effects (Fu *et al*., 2013). The 8 – 12 nucleotides directly upstream of the PAM (the 'seed' sequence) were the most important in determining sequence targeting, meaning that SNPs in the PAM distal region may not discourage off-target effects (Jinek *et al*., 2012; Cong *et al*., 2013; Hsu *et al*., 2013; Jiang *et al*., 2015; Fu *et al*., 2013; Pattanayak *et al*., 2013). Mismatched sequences that have either a base missing (gRNA bulge) or an additional base (DNA bulge) in the genomic DNA are also subject to off-target effects (Lin *et al*., 2014). Increasing the length of the gRNA can also increase specificity, with the addition of two bases at the 5′ end improving the specificity of the system but reducing the activity in human cells (Cho *et al*., 2014). On the other hand, shortening the gRNA target sequence to 17 nucleotides reduced off-target effects without affecting the mutational activity at target sites (Fu *et al*., 2014).

High concentrations of Cas9 nuclease and gRNAs can increase off-target effects so controlling *Cas9* expression could increase specificity (Hsu *et al*., 2013; Pattanayak *et al*., 2013). In the ZFN and TALEN systems, the associated nuclease (*Fok*1) functions as a dimer, with each monomer in the dimer cleaving a single DNA strand (reviewed in Bortesi and Fischer, 2015). The Cas9 system can be modified to be similar, with the D10A mutation in the RuvC nuclease domain converting the Cas9 to a nickase. A pair of these nickases can be used, increasing the specificity of the system, as two target sites would have to be present to create the DSB and for editing to occur. This method has been employed in human and mouse cells (Cho *et al*., 2014; Mali *et al*., 2013b; Ran *et al*., 2013; Shen *et al*., 2014b) and Arabidopsis (Fauser *et al*., 2014; Schiml *et al*., 2014), with efficiencies comparable to unmodified Cas9 but with greater specificity. This method does require the design of two equally efficient gRNAs, which limits target sites, and gRNAs are known to differ in their targeting efficiencies (Cho *et al*., 2014). Catalytically inactive Cas9 nucleases have also been fused to *Fok*I nuclease, to produce a system which is comparable to the nickases in activity but, again, with greater specificity (Guilinger *et al*., 2014; Tsai *et al*., 2014).

4.1.6. Protoplast system

Protoplast systems have been used previously for the generation of edited plants, with a published rice-optimised protocol available since 2014 (Shan *et al*., 2014). Wheat plants cannot be regenerated

from protoplasts; however, the development of a wheat-optimised protoplast assay system for the trial of potential gRNAs for targeting the Cas9 nuclease could reduce the temporal and financial costs of the CRISPR/Cas9 editing system. Wheat is recalcitrant to transformation and the generation of transgenic plants is an expensive and time-consuming process. Further to this, gRNAs are unreliable and sometimes do not work despite stringent selection processes (Thyme *et al*., 2016). The generation of a transient protoplast system would allow multiple gRNAs to be trialled and evaluated before stable transformation was attempted. The efficiency of different gRNAs could be assessed, and successful gRNAs selected for stable transformation attempts, either through biolistic bombardment or through an Agrobacterium-based system.

4.1.7. CRISPR/Cas9 edit detection and characterisation in transformed plants

One of the main problems with genome editing is the detection of the editing events that are produced. The editing efficiency of the CRISPR/Cas9 system can be verified using reporter genes, such as those encoding β-glucuronidase or a fluorescent protein (GFP, YFP, or RFP). The reporter gene can be designed to contain a target site that causes a frameshift, with successful mutation of the target by the Cas9/gRNA complex correcting the reading frame of the gene, and restoring function (Jiang *et al*., 2013b; Feng *et al*., 2014). The reporter gene could also contain a duplicated region, and DSBs introduced by the CRISPR/Cas9 system could lead to recombination between the duplicated regions, restoring gene function (Siebert and Puchta, 2002; Mao *et al*., 2013). However, whilst useful in testing the efficacy of the system, this is not particularly useful for detecting edits when the target is a native gene.

Polyacrylamide gel electrophoresis (PAGE) can be used to identify mutations through single-strand conformation polymorphism (SSCP), as single-stranded DNA molecules with nucleotide variations can have conformational changes and thus have different migration rates (Zheng *et al*., 2016b). This method can also detect heteroduplex DNAs with targeted mutations (Zhu *et al*., 2014). Nucleotide changes in DNA strands can also affect their melting temperature, and so high-resolution melting can be used to detect differences in PCR amplicons (Dahlem *et al*., 2012; Fauser *et al*., 2014); however, the detection sensitivity is relatively low and gives no information on the specific edits present.

Another frequently used detection method for editing events is PCR restriction enzyme digestion assay (PCR-RE), whereby the target genes contain restriction enzyme sites close to the PAM sequences, so that the restriction sites may be disrupted when successfully targeted by the Cas9 nuclease. Restriction Fragment Length Polymorphism (RFLP) is then used to identify samples containing edits in that target region, whereby the failure of restriction enzyme digestion indicates the occurrence of DNA sequence mutations. A restriction digest before or after the PCR amplification can enrich for edited sequences (Voytas, 2013; Jiang *et al*., 2013b; Nekrasov *et al*., 2013; Shan *et al*., 2013).

The Surveyor nuclease and T7 Endonuclease I assays are two methods that allow for the detection of mutations at a specific locus (Mao *et al*., 2013; Xie and Yang, 2013; Wang *et al*., 2014). They are enzymatic mismatch cleavage assays that target hybrid DNA fragments in PCR amplicons. When allowed to anneal, heteroduplexes form between wildtype and edited sequences, and if that DNA duplex contains unpaired nucleotides it will be cleaved by the nuclease (Voytas, 2013; Vouillot *et al*., 2015). This method has the benefit that it can be applied to any target sequence, but the detection sensitivity is lower than for other methods (Voytas, 2013; Vouillot *et al*., 2015).

Deep (high-throughput) sequencing of a whole genome, or of specific PCR amplicons, can provide a wealth of information on the mutations present in an organism, and can detect rare (low frequency) mutations as well as complicated chimeric mutations, making it particularly useful in assessing possible off-target mutations over the whole genome (Fauser *et al*., 2014; Feng *et al*., 2014). Target sites can be amplified by PCR, cloned, and their nucleotide sequences obtained by Sanger sequencing of multiple clones. This method can detect simple or complicated mutation patterns, and the use of restriction sites, if present, can enrich for the mutant DNA (Shan *et al*., 2013); however, this method is generally expensive and time-consuming. If the CRISPR/Cas9 system generates a high proportion of uniform mutations (including biallelic, homozygous, or heterozygous mutations) in the T0 plants, such as has been seen in rice (Zhang *et al*., 2014a; Zhou *et al*., 2014; Ma *et al*., 2015b), the cloning step might not be needed. This direct sequencing of PCR amplicons can result in superimposed sequencing chromatograms if two different edited alleles are present, which can be difficult to interpret. Thus, tools such as the DSDecode (http://dsdecode.scgene.com/) (Liu *et al*., 2015b) have been developed to decode the overlapping chromatogram by a method called degenerate sequence decoding (DSD) (Ma *et al*., 2015a). Nucleotide sequencing methods for edit detection are further discussed in Chapter 6.

4.1.8. Chapter aims

The aim of the work described in this chapter was to design a series of gRNAs to edit the *TaASN2* gene of wheat in order to reduce TaASN2 protein function. The first exon of *TaASN2* was targeted because it encodes the glutamine binding domain of the enzyme. Four gRNAs, gRNA1-4, were designed to target sites in the A, B and D genomes simultaneously. A plasmid multiplexed construct, based on the methods outlined by Xie *et al*. (2015), containing the four gRNAs, was successfully designed.

Wheat protoplasts were generated in the hope of constructing a protoplast trial system for gRNA selection, prior to the stable transformation of wheat embryos.

4.2. Results

4.2.1. Designing gRNAs

Four gRNA guide sequences were designed to target the first exon of *TaASN2* (Table 4.1). The gRNAs were designed to target regions that are conserved between the genomes, to ensure that all homeologues were simultaneously targeted for editing, but that are not present in the other *TaASN* genes. The first exon of *TaASN2* was targeted as it encodes the glutamine amidotransferase domain of the protein, which is involved in the binding of glutamine and the removal of an amino group to produce glutamate and ammonia. The gRNAs were initially designed with the aim of causing a large (244 bp) deletion between the outer gRNAs (gRNA1 and gRNA4 in Figure 4.1), removing the part of the gene encoding a large section of the glutamine binding domain and bringing about a frameshift in the coding sequence.

Table 4.1. Guide RNAs designed to target *TaASN2*, shown with the PAM sequences underlined. Scores in red indicate sub-par scores dictated by the gRNA generator, according to their scoring algorithms. RGEN scores provide the GC content and an 'out-of-frame score', where a low score indicates a higher predicted level of off-target activity. DESKGEN scores give an estimation of the gRNA activity, where a low score represents low predicted activity, and the potential for off-target effects, where a low score denotes a higher number of potential off-target hits. Scores of less than 40 in the activity and less than 60 in the off-target scores in the DESKGEN scores, indicate sub-par scores. The 3D structure is scored on whether the gRNA bases are unlikely to form stem loops and therefore are 'free' to interact with the wheat DNA.

Name	Sequence (with PAM)	Direction	RGEN scores (GC, out-of-frame)	DESKGEN scores (activity, off target)	3D structure
gRNA1	GGGGTGCGGCGACGAGTCGCAGG	Forward	80, 66.1	35, 80	Passable
gRNA2	GGACTGGAGCGGCCTGCACCAGG	Forward	75, 51.3	48, 87	Passable
gRNA3	GTAGAGCGGCTGGTCGCCGGAGG	Reverse	75, 58.7	60, 40	Good
gRNA4	CCTCGCAGTCACTGCCGGTCCGG	Reverse	70, 68.5	47, 49	Best 3D structure

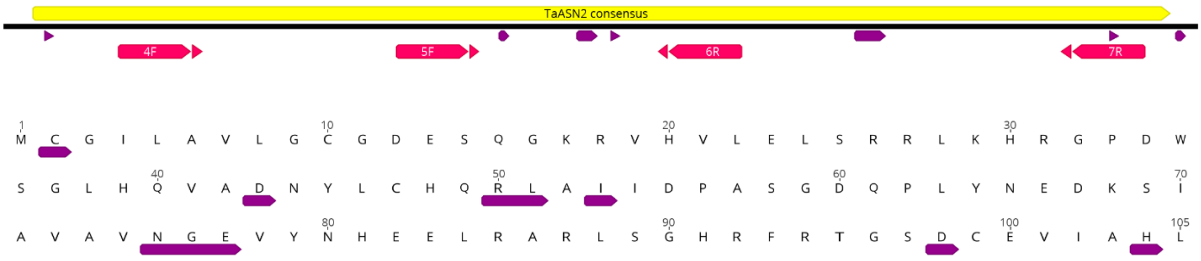


Figure 4.1. Top: The position of the gRNAs (shown in pink) designed to target the glutamine binding domain in the first exon of *TaASN2*, highlighted in yellow. Bottom: the amino acid sequence encoded by the first exon of *TaASN2*, with the residues important in the glutamine binding domain shown in purple.

The potential gRNAs (Table 4.1; Figure 4.1) were generated through the RGEN program (Park *et al.*, 2015), and DESKGEN (Hough *et al*., 2016). For the RGEN program, the '*Hordeum vulgare* (Ensembl Plants 28)' genome was used as the target genome because a wheat reference genome was not available at the time; however, the raw target sequence used was the A homeologue of *TaASN2*. The gRNAs target a region of the gene that is 40, 115, 175 and 284 bases downstream of the start of translation (gRNA1-4, respectively). Both gRNA1 and gRNA2 are oriented in the forwards direction, whilst the third and fourth gRNAs are in the reverse orientation. They were aligned to the *TaASN2* Cadenza alignment (using the Geneious software package), to confirm that they would hit all the homeologues. As only the *TaASN2* genes were to be targeted, the gRNAs were then aligned to the other *TaASN* genes, particularly the *TaASN1* genes (Figure 4.2). This was to assess the likelihood of potential off-target effects in the other *TaASN* genes, due to the high-level of identity between the genes in the family. While the chosen gRNAs correspond to regions of shared identity between the *TaASN2* genes, they all show at least two SNPs to the *TaASN1* genes, with the gRNA4 position also lacking the requisite PAM sequence in the *TaASN1* genes.

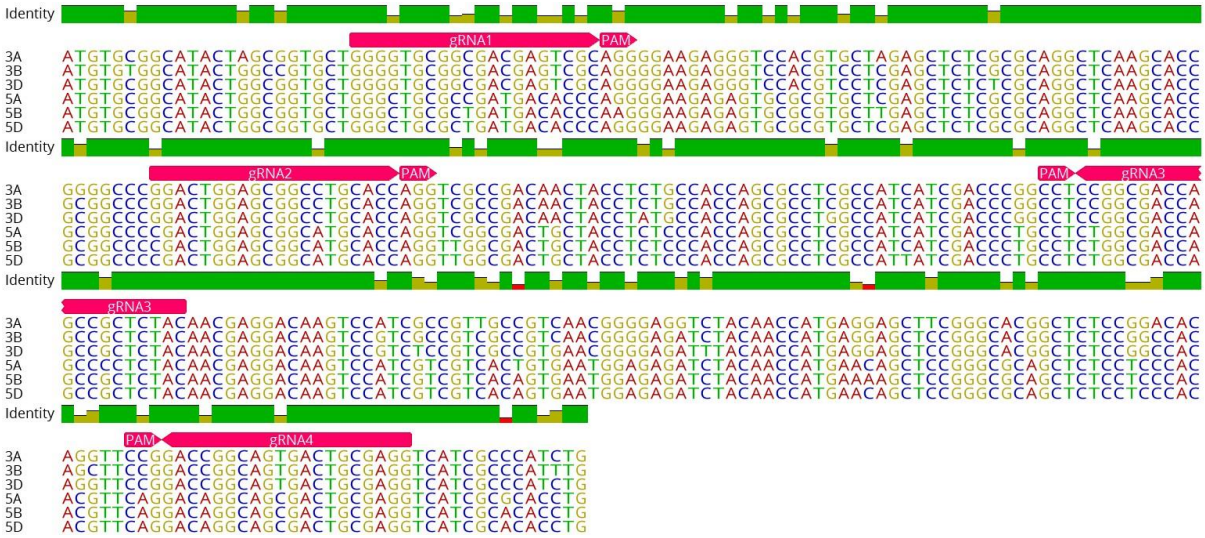
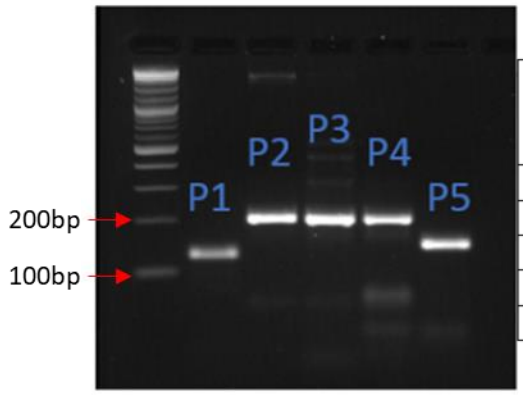


Figure 4.2. Alignment of the first exon of the *TaASN1* (5A-5D) and *TaASN2* genes, with the chosen gRNA positions highlighted in pink. The identity bar shows the positions of SNPs between the sequences. The alignment was generated using the Geneious software package (version 10.1.3).

The gRNAs target the gene at positions 40, 115, 175 and 284 bases downstream of the start of translation (in order, gRNA1-4). Both gRNA1 and gRNA2 are oriented in the forwards direction, while the third and fourth gRNAs are in the reverse orientation.

4.2.2. Multiplexed gRNA plasmid generation

A gRNA construct containing a 4-gRNA polycistronic gene was generated *via* a Golden Gate (GG) assembly to make plasmid pRRes209.481.ASN2, based on the multiplexed system proposed by Xie *et al*. (2015). The polycistronic 4-gRNA structure consisted of gRNAs (20 bp guide sequences + gRNA scaffold) interspaced with tRNAs, and was generated by the GG assembly of overlapping PCR fragments. This 4-gRNA fragment was then ligated into the backbone plasmid, pRRes208.481. The gRNAs incorporated an optimised gRNA scaffold structure (Dang *et al*., 2015). A PCR reaction was set up using specially designed primers (Chapter 2; Section 2.3.6.) to generate the required fragments for the generation of the 4-gRNA polycistronic gene (Figure 4.3).



Fragment	Forward Primer	Reverse Primer	Expected Product Size (bp)
P1	L5AD5-F	gRNA1-R	129
P2	gRNA1-F	gRNA2-R	205
P3	gRNA2-F	gRNA3-R	205
P4	gRNA3-F	gRNA4-R	205
P5	gRNA4-F	L3AD5-R	148

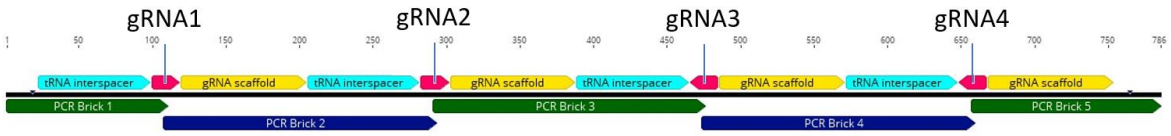


Figure 4.3. Top left: Electrophoresis gel showing level 1 fragments generated by PCR amplification. The PCR product was run against the 2-Log DNA Ladder (0.1-10.0 kbp) (New England Biolabs, Hitchin, UK). Top right: The primer combinations used to generate the PCR fragments and their expected product size. Bottom: diagram of the designed fragments (PCR Bricks 1-5) containing the gRNA, gRNA scaffold and tRNA sequences.

The PCR products were then extracted and quantified (concentrations ranged from 112.3-179.1 ng/μL). The fragments were successfully combined in a GG reaction, reamplified using generic primers S5AD5-F and S3AD5-R (Chapter 2; Table 2.3), leading to the generation of a 4-gRNA fragment of the expected size of 804 bp (Figure 4.4). Problems occurred during the re-amplification of the GG-generated gRNA multiplex, where the PCR reaction produced a ladder of bands. This was unlikely to be due to primer non-specificity, as attempting to optimise the reaction seemed to have no effect. Further to this, similar results have been seen when trying to amplify a 7-gRNA fragment, containing gRNAs designed to target genes in *Triticum turgidum* cv. Svavo (Camerlengo *et al*., 2020).

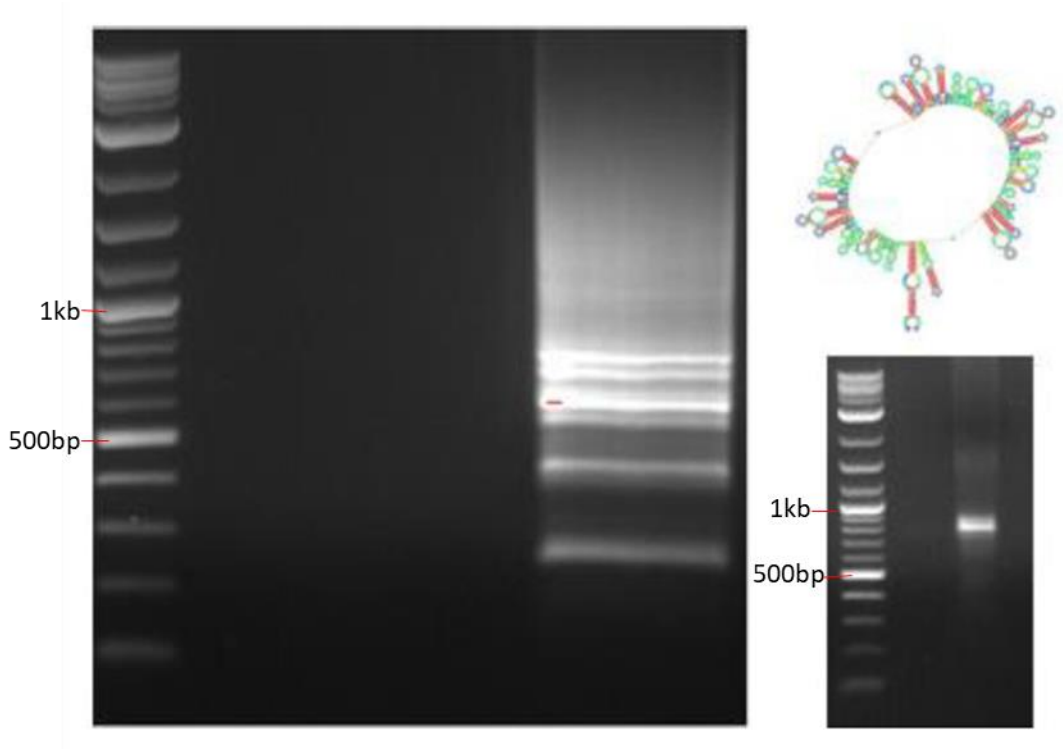


Figure 4.4. Left: Electrophoresis gel showing the 4-gRNA Level 1 fragment, reamplified by PCR. A ladder of bands is shown, and the top band on the gel corresponds to the band of interest at 804 bp. The product was run against the 2-Log DNA Ladder (0.1-10.0 kbp) (New England Biolabs, Hitchin, UK). Top right: A diagrammatic representation of the predicted secondary structure of the Level 1 fragment, generated using the Geneious software package (version 10.1.3). Bottom right: Electrophoresis gel showing the *Fok*1-digested Level 1 fragment, estimated at a concentration of 10 ng/μL.

The predicted secondary structure of the PCR amplicon (Figure 4.4, right) shows the formation of multiple 'loops' in the structure. Both the gRNA scaffolds and the tRNA interspacers have complex secondary structures. These loops may cause the polymerase to 'jump' during replication, leading to multiple smaller bands where loops have been missed out. This would result in a mix of PCR fragments of varying lengths and could explain the pattern seen on the gel. This could have been confirmed by nucleotide sequence analysis of all the bands; however, a band of the correct size was identified, and this was extracted from the gel, purified and quantified (29.9 ng/μL). This extracted band was digested with *Fok*1, and added to a backbone pRRes209.482 plasmid pre-digested with *Bsa*1, at two different ligation ratios (vector:insert = 5:3 and 5:2). The concentration was estimated at around 10 ng/μL from gel electrophoresis (Figure 4.4, bottom right).

The ligation mixture was used to transform *E. coli* with the plasmid backbone (pRRes209.481) used as a control. Four colonies were chosen to generate DNA minipreps (the samples were quantified, and concentrations ranged from 51.1 to 83.5 ng/μL) and a diagnostic digest with *Mlu*1 and *Pst*1 was run

(Figure 4.5). Two of the samples, ligation 1 colony 1 and ligation 2 colony 2, were sent for nucleotide sequencing and confirmed as the desired plasmid (pRRes209.481.ASN2). Larger amounts of the construct were generated for the purposes of transformation of both *Triticum aestivum* cv. Cadenza embryos and leaf protoplasts.

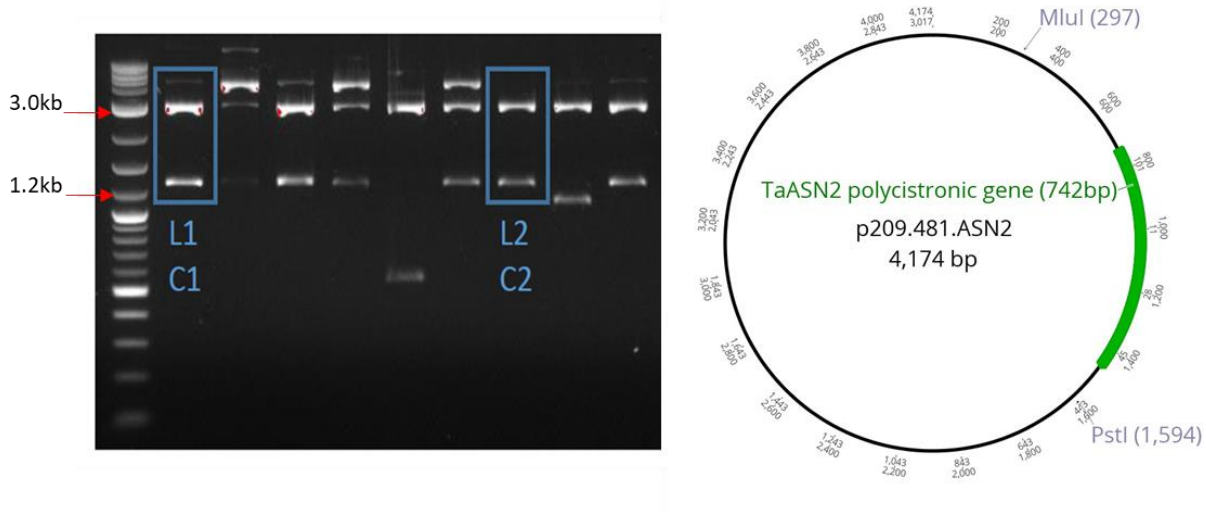


Figure 4.5. Electrophoresis gel with plasmid DNA extracted from *E. coli* colonies transformed with the ligation mix of the Level 1 GG assembly and digested with *Mlu*1 and *Pst*1. The first four lanes correspond to a transformation vector-to-insert ratio of 5:3, the fifth lane corresponds to the control plasmid pRRes209.481, and the remaining four lanes correspond to a ligation ratio of 5:2. Right: Diagram representing the structure of plasmid pRRes209.481.ASN2 with the restriction sites.

4.2.3. Generating the Actin-Cas9 plasmid

The usual *Cas9*-containing construct used at the University of Bristol placed the *Cas9* gene behind a ubiquitin promoter (Ubi-Cas9). For this project, the *Cas9* gene was placed behind an *Actin* gene promoter, *via* a two-step digestion, using the enzymes *Spe*I and *Eco*RI, to ligate the *Cas9* fragment (from the *Ubi*-driven *Cas9* construct) into the Actin-eGFP backbone plasmid, pRRes208.381. This would provide more options for *Cas9*-construct choices, as well as providing the opportunity to co-transform with the two different *Cas9* constructs.

The fragments were ligated together, and the ligation mixes were used to transform *E. coli* cells. Five different ligation mixes were set up: 1) vector and buffer, no insert or ligase; 2) vector, buffer and ligase, no insert; 3) vector, buffer and insert, no ligase; 4) all components; and 5) all the components, with double the amount of the insert. Colonies were present on ligation mix 2 (12 colonies), mix 4 (3 colonies) and mix 5 (3 colonies).

PCR was performed, using the Actin-1F or -2F primers with the Cas9-1R primer, on colonies from ligation mixes 2, 4 and 5 to confirm the presence of the *Cas9* gene in the vector. Colonies 1, 2 and 3

were obtained from ligation mix 4; colonies 4, 5 and 6 were obtained from ligation mix 5; and colonies 7 and 8 were obtained from ligation mix 2 (which did not have the insert added). No bands were seen for the control colonies used. This suggests that the 12 colonies seen on the plates were the product of self-ligation of the pRRes208.381 plasmid, as was presumed. Successful amplification was obtained for colonies 2, 4, 5 and 6 (Figure 4.6).

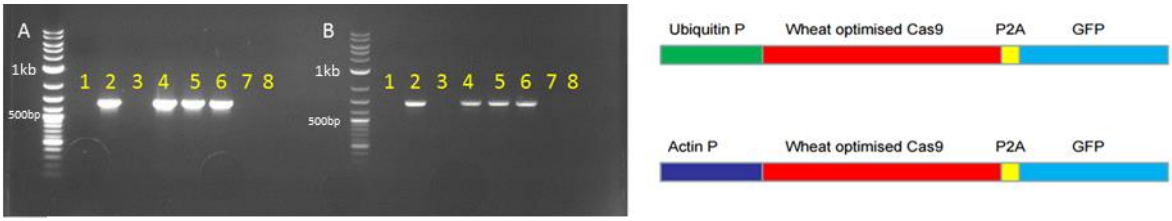


Figure 4.6. Left: Electrophoresis gel showing PCR products from colonies using either: A) Actin-1F or B) Actin-2F, in conjunction with Cas9-1R primer. Bands demonstrate successful ligation through the presence of the *Cas9* insert. Right: diagram of the two different constructs.

Plasmids were extracted from LB cultures inoculated with colony 2, 4, 5 or 7 (no PCR amplification observed), and the nucleotide sequences were analysed using the Actin-1F, -2F and Cas9-1R primers. Colonies 4 and 5 were found to contain the correct plasmid.

4.2.4. Protoplast generation

A gRNA-assessment system was developed in wheat protoplasts through the transient expression of transgenic constructs. Wheat leaf cells were harvested to generate wheat protoplasts, and these were transformed with a construct containing a *Cas9* gene and at least one gRNA. The transformation efficiency was assessed by the inclusion of a nuclear-tagged GFP in the designed constructs (Figure 4.6 Right; Figure 4.7) and the protoplast DNA was extracted. The target region was amplified by PCR and assessed by gel electrophoresis and nucleotide sequencing to identify potential editing events. Protoplast transformation was increased through protocol optimisation, to around 70 % in some cases; however, larger plasmids, such as plasmids carrying the *Cas9* gene, had lower transformation efficiencies.

The two promoters, the wheat *Actin* gene promoter and the rice Ubiquitin *(Ubi*) promoter, were compared for efficiency by the transformation of protoplasts with two constructs: pRRes208.381 (Actin-eGFP) and the ubiquitin-driven eGFP plasmid; these plasmids differ only by the promoter used. It was hoped that using a different promoter might increase the transformation efficiency seen in protoplasts and thus increase *Cas9* activity and the likelihood of editing events. However, the results of the promotor comparisons were inconclusive. Both promoters worked with varying efficiencies, dependent on the different optimisation methods used and different protoplast batches (*Ubi*: 50.44 – 61.23 %; *Actin*: 31.60 – 65.61 %). Insufficient data were collected to generate any firm conclusions

on differences in promoter efficiency. Subsequently, optimisation of the protoplast protocol led to increased transformation efficiencies in both promoter types.

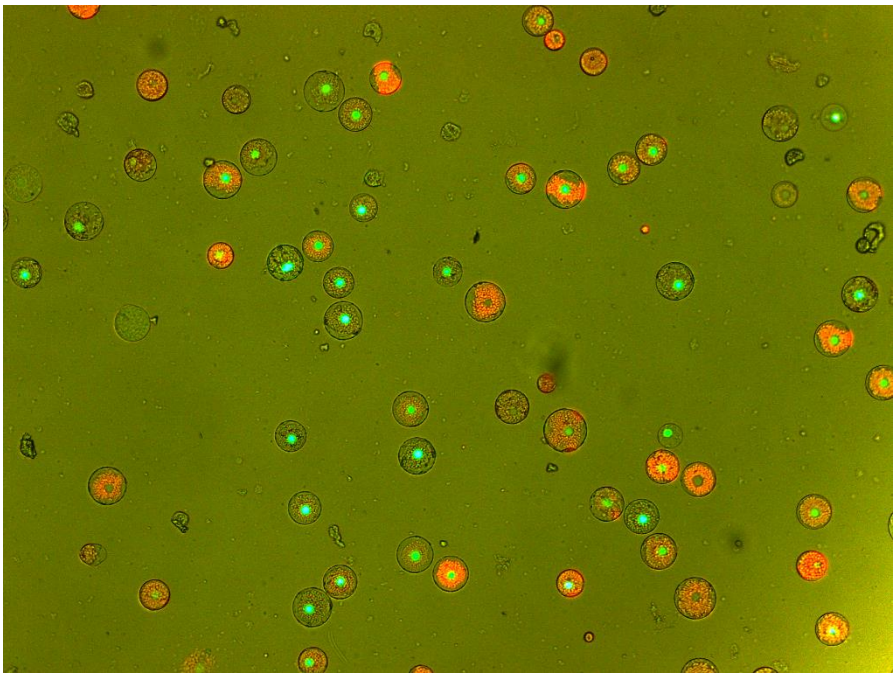


Figure 4.7. Wheat protoplasts (cv. Cadenza) 24-hours after transformation with a ubiquitin-driven eGFP construct containing a Histone2B-tagged *eGFP*, resulting in nuclear-targeted eGFP fluorescence. The protoplasts were generated using a wheat-adapted version of the Shen *et al*., (2014a) protocol. The red fluorescence is the fluorescence of the chloroplasts. Transformation efficiency was assessed from the ratio of transformed to untransformed protoplasts.

4.2.5. Protoplast trial assays

Two gRNAs, 'gRNA1' and 'gRNA2', were used to transform wheat protoplasts (pEXA24F-CONSENSUS.sgRNA; pEXA25F-CONSENSUS.sgRNA) and the editing success was estimated from protoplast DNA which was extracted 48 hours after transformation. Four different transformations were set up: 1) both gRNA constructs but no *Cas9* construct; 2) both gRNA constructs in conjunction with the Actin-Cas9 construct; 3) both gRNA constructs in conjunction with the Ubiquitin-Cas9 construct; and 4) both gRNA constructs and both the actin and the ubiquitin constructs. The target region was successfully amplified from the extracted DNA using the ASN2-1F and ASN-1R primers. Unfortunately, no large deletion was seen. Had both gRNAs successfully guided *Cas9* to the target sites, there was the possibility that editing would cause the release of the 75 bp fragment between the

two gRNAs. The PCR amplicon was purified and sequenced to investigate other potential editing events, such as the addition or deletion of a few bases around the target site.

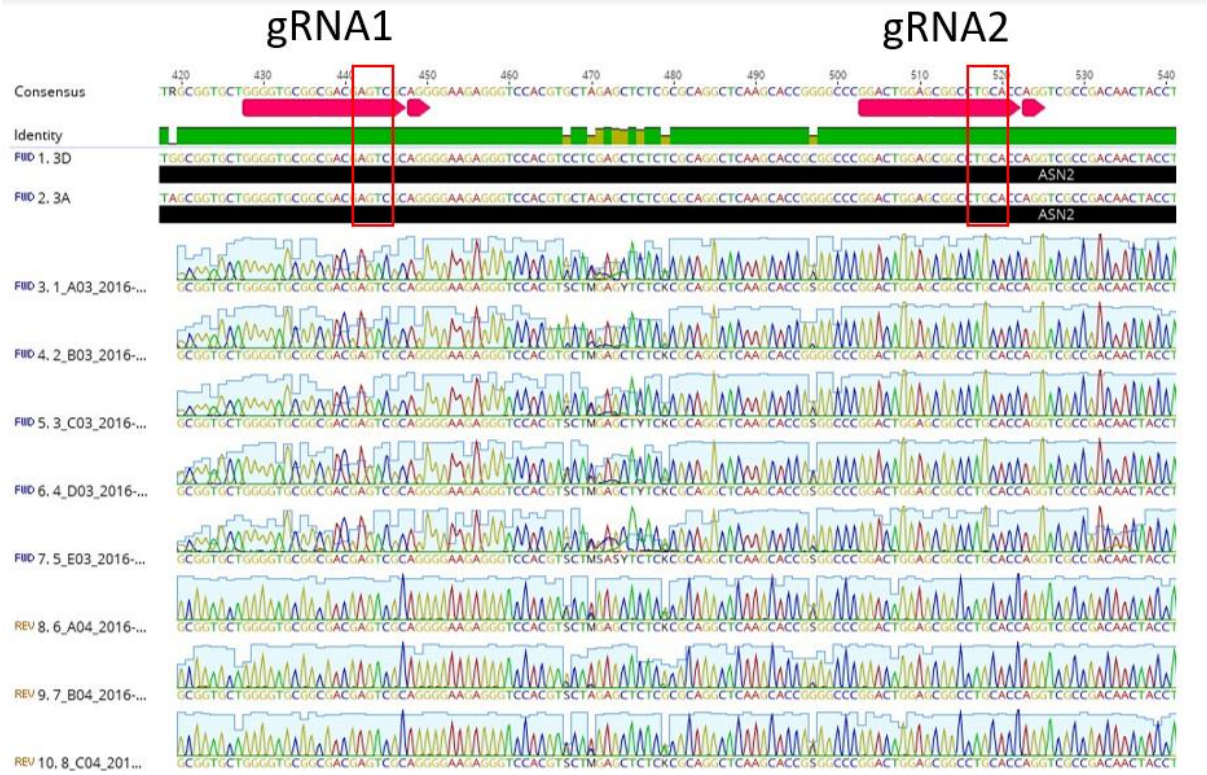


Figure 4.8. Nucleotide sequence analysis of the *TaASN2* gene in extracted DNA of Cadenza protoplasts after transformation with pRRes208.381+Cas9, Ubi-Cas9, and the 'pEXA24F-CONSENSUS.sgRNA and pEXA25F-CONSENSUS.sgRNA constructs. The nucleotide sequence analysis was conducted on a capillary sequencer; the first 5 chromatograms represent reactions primed with the ASN2-1F primer and the last 5 chromatograms represent reactions primed with the ASN2-1R primer. The red boxes highlight the expected Cas9 cleavage position.

Whilst the chromatograms show a region in between the two gRNAs that shows decreased identity between the sequencing reads and the reference sequences (Figure 4.8), this was not taken as evidence of editing events. No mutations or deletions were seen at the Cas9 cutting positions, which is taken as 3 bases upstream from the PAM sequence (Figure 4.8, red boxes). The reads were generated from a population of extracted protoplast DNA, rather than a single sample. It is therefore possible that editing events did occur but were not detectable using this system as they would account for a small proportion of the total population.

The gRNA2 target sequence overlapped the *Sex*A1 restriction enzyme recognition site, where successful editing would disrupt the recognition site. Thus, if editing events had taken place, a larger, un-cut band would be present on the gel alongside the smaller, cut bands (Figure 4.9). The frequency

of editing in the whole population was likely to be very low; therefore, the un-cut band would be seen faintly in conjunction with the normal digested band pattern. This larger band was only seen in the transformation mix which contained both versions of the Cas9 construct (1. Ubi-Cas9 and 2. Act-Cas9) and the gRNAs (Figure 4.9).

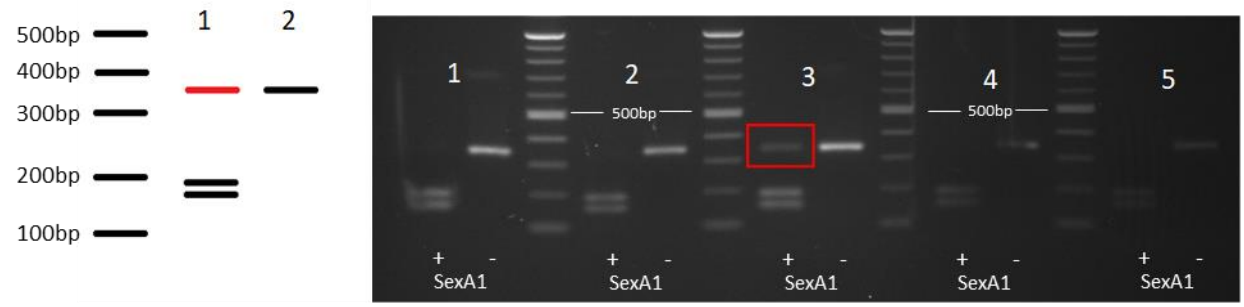


Figure 4.9. Electrophoresis gel showing digestion of the PCR amplicon with *Sex*A1. Left) The expected band pattern for 1) digest with *Sex*A1, where the red band represents the uncut fragment, and 2) the fragment with no *Sex*A1 added. Right) *Sex*A1 digest for the transformation mixes. The lanes are ordered *Sex*A1, no *Sex*A1. 1) Ubi-Cas9 and gRNAs; 2) Actin-Cas9 and gRNAs; 3) Ubi-Cas9, Actin-Cas9 and gRNAs, an uncut band is present (highlighted in red); 4) gRNAs only; and 5) Ubi-Cas9 and Actin-Cas9 only.

Although there was believed to be no difference in the two constructs, there was potentially an interesting dosage or cumulative effect of the two Cas9 constructs in tandem. The undigested pattern was only observed in the transformation where both Cas9 versions were used. No undigested bands were present in the control transformation (transformation 4, no Cas9 was added; transformation 5, no gRNAs were added).

The undigested band was purified from the gel to focus only on those sequences that had been potentially edited by gRNA2. This population was analysed to identify potential editing patterns (Figure 4.10).

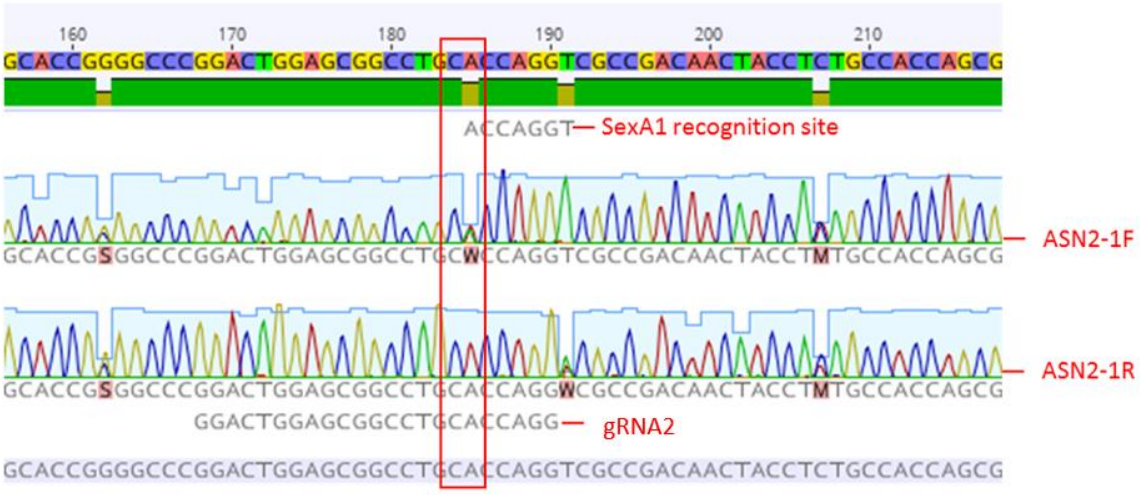


Figure 4.10. Nucleotide sequence analysis of the *Sex*A1-undigested band, from a capillary sequencer, visualised using the Geneious software package (version 10.1.3). The top chromatogram shows the read from the ASN2-1F primer and the bottom chromatogram shows the read from the ASN2-1R primer. The position of the gRNA2 is highlighted, and the red box represents the expected editing site.

The results, whilst showing that some editing may have taken place, were not conclusive. An ambiguity code was shown at the expected editing site (Figure 4.10, red box); however, its presence was dependent on whether the forward or reverse primer was used. Since the non-edited original base would be an 'A' and the call was a 'W' (either an A or T), it was not possible to be confident that editing had occurred. Further to this, no 'indel-signature' was seen. The sequence reads after the expected editing sites showed high identity and a high call confidence. Had an indel been present, it would cause a mixed chromatogram with ambiguity bases, where a deletion or insertion caused a frameshift and a reduction in sequence identity. This produces a distinctive 'messy' call signature in the chromatogram, which was not seen here.

4.2.6. Primer design for target amplification

Primers were designed to allow for amplification of the target fragment, the first exon of *TaASN2*, in Cadenza. Whilst ASN2-1R corresponded to a conserved region in the Cadenza sequence, ASN2-1F did not. Thus, two new forward primers (a new ASN2-1F, and ASN2-2F) and a new reverse primer (ASN2-2R) were designed. The issue in designing the primers for the first exon of *TaASN2* is the difference in the intron length immediately following the exon (the first intron), reducing areas of homeologue identity (Figure 4.11).

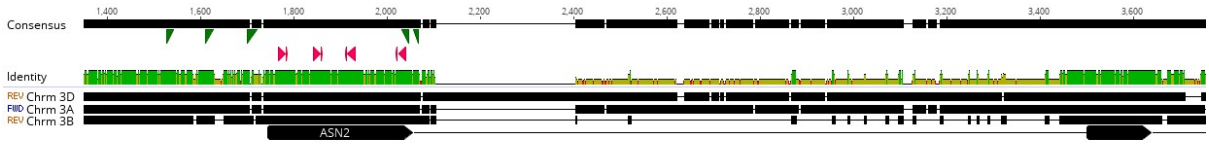


Figure 4.11. Diagram illustrating the deletion in intron 1 of chromosome 3A (302 bp) and chromosome 3B (1236 bp). The exon-intron structure is shown by annotations. Exons are denoted by the black arrows and introns are denoted by the lines connecting them.

Chromosome 3B possesses a deletion of about 1250 bp in the first intron of *TaASN2*, directly after the 5' splice site (Figure 4.11). To design effective primers for this system, fully conserved (non-specific) primers were required which would target all the homeologues simultaneously. Furthermore, the fourth gRNA (gRNA4) sits close to the end of the first exon of *TaASN2*. Reactions were set up with the primers in four different pairs and trialled at eight different temperatures (55-70 °C), with the expected product sizes being between 440-543 bp (Figure 4.12).

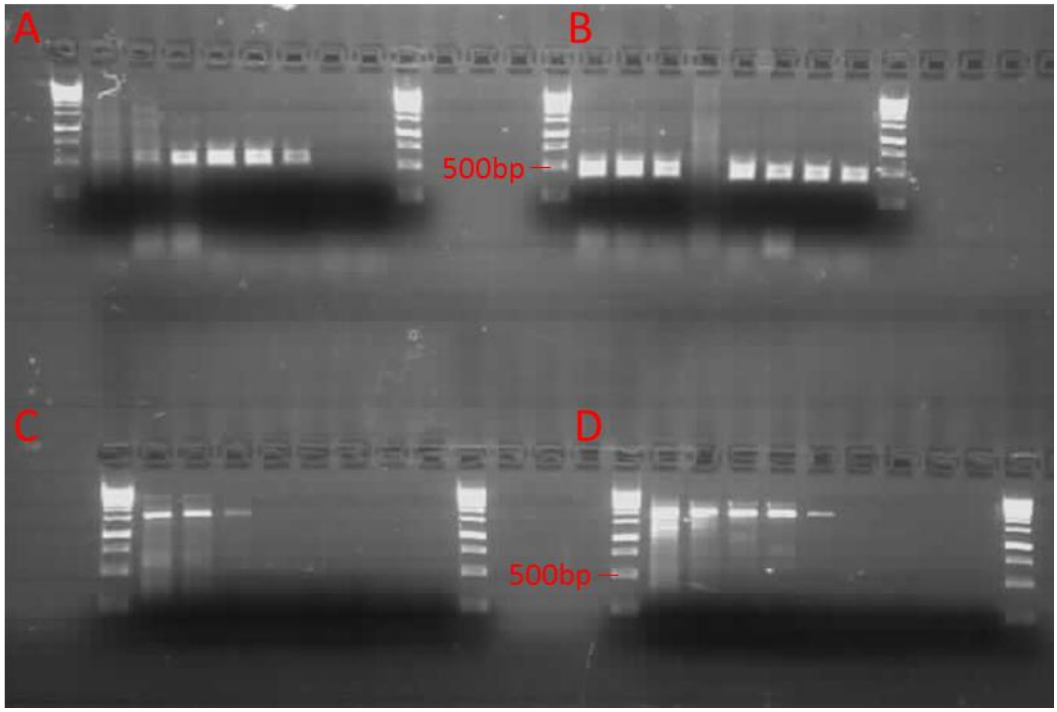


Figure 4.12. Electrophoresis gel showing the products of a PCR optimisation experiment for the potential ASN2-1F, ASN2-2F, ASN2-1R and ASN2-2R primer pairs. A = ASN2-1F and ASN2-1R, with an expected product size of 524 bp; B = ASN2-2F and ASN2-1R, with an expected product size of 440 bp; C = ASN2-1F and ASN2-2R, with an expected product size of 543 bp; D = ASN2-2F and ASN2-2R, with an expected product size of 459 bp. The PCR reactions had been run on a temperature gradient, left to right (all in °C): 55, 56, 58, 60.9, 64.5, 67.5, 69.2, and 70.

Bands of the expected size were only seen for reactions A and B, which both involved the ASN2-1R primer (Figure 4.12), although no band was observed in the fourth well (60.9 °C) of reaction set B. The presence of larger bands for reactions C and D suggested non-specific amplification.

4.2.7. Protoplast transformations for the trial of the construct pRRes209.481.ASN2

Wheat (*Triticum aestivum*) cv. Cadenza protoplasts were transformed with the 4-gRNA construct, pRRes209.481.ASN2 (2.16 µg/µL), and the *Cas9*-containing pRRes217.486 (2.02 µg/µL). The transformation efficiency from the control transformations (pUbi::H2B-GFP) was estimated at 38.2 % from the eGFP fluorescent counts (taken 48-hours after transformation). DNA was extracted from the protoplasts after 48 hours (0.7 ng/µL). A PCR was run, using the primer pair ASN2-F2 and ASN2-1R, to amplify the first exon of *TaASN2*. The results were visualised on an agarose gel, where 'band-shifts' were seen (Figure 4.13).

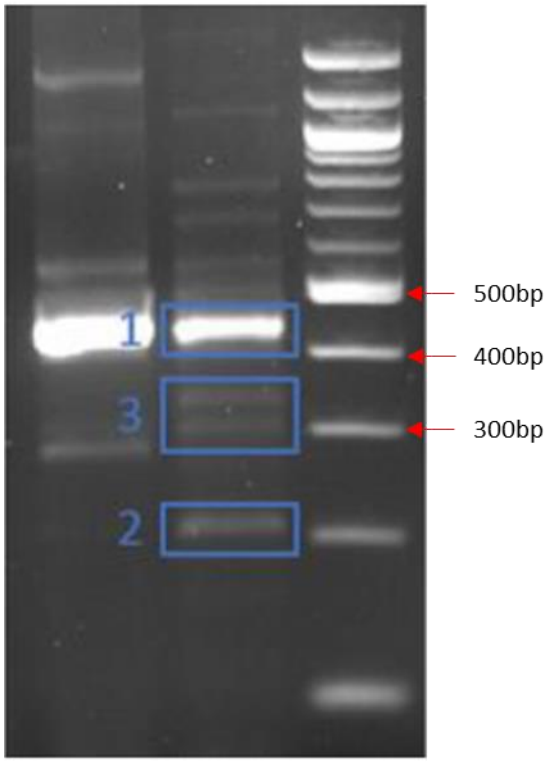


Figure 4.13. Electrophoresis gel showing PCR products generated from the first exon of *TaASN2*, amplified with the primer pair ASN2-F2 and ASN2-1R, from extracted protoplast DNA. The sample was run against a control and the 100 bp DNA Ladder (New England Biolabs, Hitchin, UK). The blue boxes show the fragments, which were extracted, cloned into TOPO plasmids and analysed.

The non-edited band sizes were expected at around 430 bp (433 bp for the A and D genome, and 421 bp for the B genome). A deletion between the gRNAs would produce bands of the following sizes (based on the A genome): gRNA1 and gRNA2 = 365 bp; gRNA1 and gRNA3 = 305 bp; gRNA1 and gRNA7 = 196 bp; gRNA2 and gRNA3 = 380 bp; gRNA3 and gRNA4 = 271 bp; and gRNA3 and gRNA4 = 331 bp. The ladder of bands seen on the gel seemed to correspond to the expected band size if editing had taken place (Figure 4.14), although it appeared that gRNA2 may not have been working efficiently. The smallest band on the gel (around 200 bp) would correspond to the flanking gRNAs, gRNA1 and gRNA4, if the complete fragment between them had been excised. The bands were cut and extracted from the gel in three groups: 1) the smallest band of around 200 bp; 2) the largest fragments, which corresponded to a faint double band, at about 430 bp, which would be expected if editing had not occurred; and 3) the middle bands, of which there were at least two, ranging from 300-400 bp (Figure 4.13). The middle bands, which were impossible to cut from the gel separately, were expected to correspond to a mix of edited fragments.

The fragments were then cloned into separate plasmids containing the *ccdb* killer gene using the Zero Blunt™ TOPO™ PCR Cloning Kit (Thermo Fisher, UK). *E. coli* cells were transformed with the TOPO-mixture and individual colonies were used to inoculate 3 mL of YT broth and incubated overnight. DNA was extracted from the colonies (all samples showed concentrations between 117.5 and 243.1 ng/μL) and sent for nucleotide sequence analysis. Four colonies were chosen from Group 1, and six colonies were chosen from the remaining groups. Whilst the larger fragment of 430 bp aligned to the genomic cv. Cadenza sequence of *TaASN2* perfectly, the sequencing results of the shorter bands could not confirm editing at the expected sites. Indeed, the sequences produced little or no similarity with the genomic cv. Cadenza *TaASN2* sequence. The first assumption was that there was non-specific amplification by the ASN2-F2 and ASN2-1R primer pair. However, BLAST searches of the sequence results against the cv. Cadenza genome returned no clear matches, whereas BLAST searches using the 'non-edited' sequences returned hits to *TaASN2* on all three genomes, along with hits to *TaASN1* on chromosome 5. It was possible that the amplicons corresponded to regions on the vectors; however, this was checked, and the amplicons did not match the vectors. A further protoplast transformation experiment was conducted, and the DNA was extracted. The subsequent PCR, using the ASN2-F2 and ASN2-1R primers, did not yield a band-shift pattern, although transformation efficiency was extremely low at around 23 %, and further optimisation would have been needed to make this an effective trial assay.

4.2.8. Attempted generation of a second gRNA-containing plasmid *via* GG assembly

The gRNA construct, pRRes209.481.ASN2, containing the four gRNA guide sequences designed to target exon 1 of *TaASN2*, was generated using the gRNA scaffold suggested by Dang *et al*. (2015).

This scaffold had not been used at Rothamsted Research before, so a second construct, containing the same four gRNAs but using the conventional scaffold (Mali *et al*., 2013b), was conceived. This scaffold had been used in successful wheat genome editing experiments at Rothamsted Research (Camerlengo *et al*., 2020). Despite the construct being designed in the same way as pRRes209.481.ASN2, all attempts to make this second construct failed. The major bottleneck existed in the PCR generation of the GG Level 1 fragments, which may have been caused by the high 'GC' content of both the primer and wheat chromosomes generally (as has been seen in investigations into the *Rht* genes (Pearce *et al*., 2011)), the large overhangs of the primers themselves or the complex secondary structure of both the gRNA scaffolds and the tRNAs. The function of both the tRNAs and the gRNA scaffold relies on their complex secondary structure; however, this could hamper amplification by PCR.

4.3. Discussion

4.3.1. Differences between varieties Chinese Spring and Cadenza

Although the initial gRNAs, 'gRNA1' and 'gRNA2', were designed to target the cv. Chinese Spring *TaASN2* genomic sequence, they both corresponded to conserved regions in the cv. Cadenza genome (Appels *et al*., 2018) and thus were retained as potentially viable gRNAs for editing Cadenza. They were used in conjunction with the two newly designed, untested gRNAs, 'gRNA3' and 'gRNA4', in the creation of a multiplexed gRNA construct, pRRes209.481.ASN2, for cv. Cadenza isolated embryo transformation.

Chinese Spring only contains two homeologues of *TaASN2*, one on chromosome 3A and one on chromosome 3D, whereas Cadenza contains three homeologues, with a chromosome 3B version of *TaASN2*. As such, designing gRNAs for the Cadenza genome added an increased level of complexity due to the increased number of single nucleotide polymorphisms present between the genomes. The potential gRNA sites in the first exon of *TaASN2* were limited, increasing the reliance on the optimisation of the protocol with the existing gRNAs.

Further to this, the potential for deletion events in the first intron in wheat cv. Cadenza to cause issues with the binding of reverse primers, further limited the design of reverse primers for the amplification of the first exon target site. There is the potential to exploit this, where the same set of primers would amplify dramatically different size fragments (where the 3A and 3B would produce fragments which were larger than the 3D amplicon by 302 bp and 1236 bp, respectively), allowing for discrimination between the genomes through one PCR reaction. This is being followed up in another PhD project.

Whilst the first exon of *TaASN2* is the ideal site to target to ensure a complete knock-out (i.e. to preclude the production of a functional protein), targeting the second exon would almost certainly

generate a dysfunctional protein as well, and would have been explored as a second option should the targeted editing of the first exon of *TaASN2* failed.

4.3.2. Effectiveness of protoplasts

The use of protoplasts may provide a quick and relatively easy way to assess gRNA efficiency, without the expensive investment in wheat embryo transformation. However, wheat is recalcitrant to regeneration from protoplasts. As a result, this method would be useful only to assess gRNA efficiency, and not a viable option to produce stably edited plants.

The protoplast protocol used was based on that of Shan *et al*. (2014), which was wheat-optimised to increase transformation efficiency. A new plasmid was generated that contained the *Cas9* gene, placed behind a wheat *Actin* gene promoter, using pRRes208.381 as a donor plasmid. This plasmid was used in conjunction with the Ubi-Cas9 plasmid (from the University of Bristol), and the two gRNAs, to transform *Triticum aestivum* cv. Cadenza leaf protoplasts in a trial assay system. Extracted protoplast DNA was screened using the PCR with conserved primers, and subsequent nucleotide sequence analysis of the generated amplicons. Despite evidence to suggest a disrupted recognition site for a restriction enzyme, nucleotide sequencing results did not indicate obvious editing in the extracted protoplast DNA.

Chloroplast auto-fluorescence is a well-known problem facing plant-tissue microscopy (Krause & Weis, 1991). This auto-fluorescence, with a peak of approximately 680 nm, helps maximise photosynthetic activity through a re-absorbance mechanism. This can interfere with the eGFP fluorescence signal. This issue was addressed by using dark-grown seedlings and exposing them to as little light as possible during the protoplast generation and transformation procedures. Protoplasts have a limited lifespan of only a couple of days. Upon death, the cellular products break down, emitting further auto-fluorescence as they do so. To mitigate this, nuclear-tagged GFP was used, which generated a distinctive and concentrated nuclear fluorescence, enabling the accurate identification of transformed protoplasts.

4.3.3. Detection of editing events in protoplast trial assays

The protoplast test assays showed unreliable and low transformation efficiencies, which made it difficult to judge the effectiveness of the gRNAs. Editing may not have been seen because the gRNAs were inefficient; however, alternatively, and considering the future success of the designed gRNAs, editing may not have been seen due to the inefficient transformation and poor-quality protoplast material. The use of pENT4sgRNA4.Spo11, a control plasmid that is known to reliably cause editing in protoplasts at the University of Bristol, would have allowed an assessment of the functionality of the protoplast system. However, it was decided to proceed with the gRNA constructs that had been

produced and use them to produce genome edited wheat plants (Chapter 5) rather than persevere with the protoplast trials.

The protoplast system relies on detecting edits from a PCR performed on pooled DNA, i.e. DNA extracted from a population of protoplasts. It is possible that the four gRNAs from pRRes209.481.ASN2 produced editing when transformed into protoplasts, but that these editing events were not frequent enough to be detectable. Most previous plant studies have relied on the existence of a convenient restriction site at the editing position (Belhaj *et al*., 2013; Feng *et al*., 2013; Xing *et al*., 2014; Bortesi and Fischer, 2015; Shan *et al*., 2013; 2013b; Upadhyay *et al*., 2013; Wang *et al*., 2014), with positive editing resulting in the generation of an 'uncut' band and a distinct digestion pattern of the target amplicon. In the case of protoplast populations (Shan *et al*., 2013), this uncut band can be purified from the gel, enriching for edits at that site. Another method that has been used is the detection of a 'band-shift' after the target region is amplified by PCR (Nekrasov *et al*., 2017; Belhaj *et al*., 2013). If more than one gRNA is used, there is the potential that the region between the gRNAs will be excised from the gene when it is repaired. This would result in a PCR product that is smaller than the non-edited version, producing a 'band-shift' when visualised by gel electrophoresis. The *Sex*A1 enzyme was used to screen the protoplast population; however, this would have only provided information on one of the four editing sites chosen. The first exon of the *TaASN2* gene was amplified to look for distinctive editing patterns, such as a pattern of bands of smaller sizes, showing large deletions in the target site, as has been seen previously in tomato (Nekrasov *et al*., 2017).

The detection of edits when there are no available restriction sites has proven more challenging. Not all edits produce large deletions, and we now know more about the wide range of possible edits produced from the same gRNAs (Chapter 6). Furthermore, a large deletion is not required to impair gene function and this method of detection cannot be relied upon when only a single gRNA is used. The use of a single gRNA might be a strategy to reduce potential off-target effects, although it should be noted that this is an overly-simplistic strategy, given that there is marked variability in the off-target binding sites and DNA cleavage rates among different gRNAs (Kuscu *et al*., 2014; Tsai *et al*., 2015). In wheat, this is further compounded by polyploidy. Not only may the DNA come from a population with the potential for a variety of different edits, but each protoplast itself may be edited homozygously or heterozygously in one or more of the A, B or D genome genes. Thus, the presence of a single, heterozygous, nonsynonymous polymorphism in the B genome of a single protoplast, for example, would be almost impossible to detect when analysing the entire population. The expectation was that large deletions would be created from the use of multiple gRNAs; however, analysis of the transformed wheat plants (Chapter 5; 6) has shown that this is not a frequent occurrence and that a variety of edits could occur within a single plant, let alone within an edited population. Thus, a quick,

efficient way to detect editing events from the background WT protoplast population would be needed for the protoplast system to be effective.

The coverage and number of reads generated by next generation sequencing might allow for the better detection of edits in protoplasts than can be achieved by PCR. This was not used here but was employed extensively in the subsequent analysis of edited wheat plants (Chapter 6), partly because of the experience gained in the protoplast study.

Chapter 5

Transformation, TILLING and T0 analysis

5.1. Introduction

5.1.1. Plant transformation methods

Genome editing is a breakthrough technology that has the potential to revolutionise crop research and improvement by providing targeted, precise changes in the genome. However, a key bottleneck remains with genome editing: the need for fast and effective plant transformation methods. The basis of genetic transformation is the integration of a DNA fragment, containing a gene of interest, into the host plant cell's DNA. Transgenic plants are then regenerated from the transformed cells using tissue culture methods.

Despite nearly 40 years of development, the transformation and regeneration methods for many crop species remain difficult, and this has hampered functional genomic research in those species (Shrawat and Lörz, 2006; Hiei *et al*., 2014; Altpeter *et al*., 2016). Plant tissue culture methods require optimisation for each species and expertise for success. As a result, specialised facilities, such as the Cereal Transformation Facility at Rothamsted Research, are required. In the case of Agrobacterium (*Agrobacterium tumefaciens)*-mediated transformation, methods are restricted to specific genotypes, species or genuses (Nam *et al*., 1997), and it is common that cell types that are easily transformed, such as protoplasts, are not the easiest to regenerate. Biolistic transformation can be used on a wider range of genotypes (Altpeter *et al*., 2005); however, regeneration levels can be lower than those seen in methods such as Agrobacterium-mediated transformation. There are many factors that can alter plant regeneration and successful transformation, including microparticle parameters, the quantity of DNA and particle coating, tissue type, and osmotic treatment (Frame *et al*., 2000; Kausch *et al*., 1995; Klein *et al*., 1988; Vain *et al*., 1993; Sivamani *et al*., 2009; Altpeter *et al*., 2016).

The first genetically stable transformed plant tissues were produced in the 1980s (Zambryski *et al*., 1980; Bevan *et al*., 1983; Herrera-Estrella *et al*., 1983; Fraley *et al*., 1983). In wheat, the main methods for plant transformation are biolistics and Agrobacterium-mediated transformation using immature embryos.

5.1.1.1. *Agrobacterium tumefaciens*-mediated transformation

Agrobacterium-mediated transformation is a technique based on *Agrobacterium tumefaciens* and there are similar techniques involving related *Agrobacterium* species. *Agrobacterium tumefaciens* is a well-known soil bacterium that can cause crown gall disease in plants by introducing T-DNA (transfer DNA), located on a Ti plasmid, into the plant cell (Gelvin, 2003). It is this ability to transform plant species that has been harnessed for the purposes of plant genetic engineering.

The first successful Agrobacterium-mediated transformation of wheat took place in 1997 (Cheng *et al*., 1997); however, it remained a difficult and inefficient process in most cereal species, with low

transformation efficiencies recorded in wheat (Shrawat and Lörz, 2006; Ishida *et al*., 2015). However, Agrobacterium-mediated transformation is arguably the most popular method for transforming plants and it has clear advantages over other transformation methods. These include a low transgene copy number, as well as the stable integration of intact transgenes into the plant genome (Jones *et al*., 2005). Recently, *Agrobacterium tumefaciens* strains have been successfully utilised to transform plants with the genetic components for genome editing in cereal species (Zhang *et al*., 2019; Miao *et al*., 2013; Char *et al*., 2017; Svitashev *et al*., 2015).

5.1.1.2. Biolistic transformation

Biolistic bombardment was the method employed in this project for the generation of *TaASN2* mutant wheat plants through genome editing. The first transgenic wheat produced by direct DNA transfer using particle bombardment was generated in 1992 (Vasil *et al*., 1992), before the first use of Agrobacterium-mediated transformation in wheat. It involved the transformation of embryogenic callus tissue with the *bar* gene for phosphinothricin acetyltransferase activity and tolerance of herbicides based on phosphinothrycin (PPT).

Biolistic transformation, also known as "particle bombardment" or the "gene gun technique", was designed in Cornell University in 1987 for the specific transformation of cereal species that were recalcitrant to Agrobacterium-mediated transformation (Sanford *et al*., 1987; Sanford, 1988), although its range is not limited to cereal species. The system is not dependent on host specificity or species limitation and many different plant tissues can be used as a target, including meristem, embryos and callus (Christou, 1996; Romano *et al*., 2001; Morikawa *et al*., 1989; Twell *et al*., 1989; Reggiardo *et al*., 1991; McCabe and Martinell, 1993; Christou *et al*., 1988). The method focuses on the delivery of nucleic acids to cells by high-speed particle bombardment. The technique uses nucleic acid-coated particles propelled by a pressurized gas to transform plant tissue. It is now a common method used for the nuclear transformation of plants, but it is also particularly effective for chloroplastic or mitochondrial transformations (Bonnefoy and Fox, 2007; Daniell *et al*., 1990). As well as the generation of transgenic plants, it is often used for applications such as transient gene expression studies in which the introduced DNA does not necessarily integrate into the host genome (Taylor and Fauquet, 2002; Oard *et al*., 1990; Wang *et al*., 1988; Daniell *et al*., 1990; Kartha *et al*., 1989; Morikawa *et al*., 1989; Twell *et al*., 1989; Reggiardo *et al*., 1991).

The gene gun used in biolistic transformation consists of a high-pressure and a low-pressure chamber, separated by a diaphragm. The pressure difference causes the diaphragm to break, launching a projectile along a barrel and onto a porous screen. The projectile is coated with nucleic acid-covered microparticles for the transformation procedure. The nucleic acid is most often DNA in the form of a plasmid vector and the microparticles are usually made of gold (Sandford *et al*., 1993; Sandford, 2000; Frame *et al*., 2000). The projectile is stopped by the screen but the DNA-covered microparticles

pass through and are launched towards a plate containing the target plant tissue. As the microparticles perforate the cells, the DNA is released and in a small proportion of cases may integrate into the plant DNA.

Gold particles are often used, alongside other dense metals such as platinum and tungsten, because the probability of the microparticles entering the cells is proportional to their kinetic energy. These metals also have low cellular toxicity (Southgate *et al*., 1995; Frame *et al*., 2000; Oard *et al*., 1990). Particle bombardment has been used on a range of plant species (reviewed in Rivera *et al*., 2012), and it has been employed successfully to deliver genome-editing reagents in multiple crop plants, including wheat (Zhang *et al*., 2016b; Liang *et al*., 2017; Liang *et al*., 2018; Wang *et al*., 2014), rice (Shan *et al*., 2014; Oliva *et al*., 2019) and maize (Svitashev *et al*., 2015; 2016).

When comparing biolistic and Agrobacterium-mediated transformation in rice (Dai *et al*., 2001) and barley (Travella *et al*., 2005), the Agrobacterium-mediated method showed a higher percentage of transgenic plants containing intact copies of foreign genes, especially non-selective genes, and more stable transgene expression (Dai *et al*., 2001). Furthermore, biolistic transformation often results in a high copy number of the introduced genes, which can be particularly undesirable for selectable markers and, for genome-editing purposes, the *Cas9* gene. In both methods, the DNA randomly integrates into the plant cell's DNA, with no control on integration site, so the more copies that integrate the more chance there is of inadvertent gene silencing (Anand *et al*., 2003; Alvarez *et al*., 2000; Iyer *et al*., 2000; Demeke *et al*., 1999; Chen *et al*., 1998). The plant tissue can also suffer tissue damage, leading to low throughput.

On the other hand, biolostic transformation has the advantage of circumventing the host-range limitations that can be encountered when using *Agrobacterium tumefaciens* or other *Agrobacterium* species. Furthermore, many tissue and cell types can be transformed, the transformation protocol is relatively simple, large DNA fragments can be delivered (up to 53 kbp; Partier *et al*., 2017), no specific binary vector is required, and multiple plasmids can be co-bombarded with high levels of co-transformation (Wu *et al*., 2002). The system is also not limited to DNA delivery, leading to the possibility of bombarding cells with Cas9-gRNA complexes (also known as ribonucleoprotein complexes (RNP)) and editing plants without involving DNA or a transgene (Woo *et al*., 2015; Liang *et al*., 2017; 2018; Kim *et al*., 2017).

5.1.2. Regulations governing genome-edited plants in the European Union

Whilst genome editing provides the ability to induce targeted mutations in wheat, including the rapid targeting of all three homeologues simultaneously, the regulatory situation surrounding genome-edited plants was unclear at the start of the project and remains unclear at the time of writing. An opinion issued by the European Court of Justice (ECJ) in July 2018 (Opinion Case C-528/16) on proceedings brought by Confédération Paysanne, a French farmers' union, stated that organisms

obtained by genome editing should go through the same risk assessment and approval process that is applied to genetically modified (GM) crops. GM crop approvals are time-consuming and expensive in the EU, and many Member States ban GM crop cultivation altogether (Menz *et al*., 2020). One of the consequences of the ECJ ruling was that the genome-edited plants generated in this project had to be grown in a containment glasshouse.

Another problem at the start of the project was that the intellectual property situation surrounding the CRISPR/Cas9 technology was still being contested. It was, therefore, decided that a parallel approach should be taken that utilised a wheat TILLING (Targeting Induced Local Lesions IN Genomes) population produced by chemical mutagenesis (Krasileva *et al*., 2017), because any lines produced through that approach would face no regulatory or intellectual property barriers to commercialisation.

5.1.3. TILLING population

Chemical and radiation mutagenesis have been used in crop improvement since the late 1920s, although the genetic redundancy in wheat, due to its polyploid nature, makes it difficult to screen wheat mutant populations on the basis of phenotypic changes (Dubcovsky and Dvorak, 2007; Wang *et al*., 2014; Borrill *et al*., 2019). This is overcome by TILLING, a reverse genetics method that allows for the identification of point mutations in specific genes in a population, regardless of any phenotype. Wheat is especially well-suited for TILLING due to the high mutation densities tolerated by polyploids (Tsai *et al*., 2013; Krasileva *et al.*, 2017). Mutations in the individual homeologues can be identified and combined to generate loss-of-function mutants, which will overcome the effect of genetic redundancy.

Two TILLING populations were available, a tetraploid durum wheat (*Triticum durum*) cv. 'Kronos' population and a hexaploid bread wheat cv. 'Cadenza' population (Krasileva *et al.*, 2017). They had been developed as part of a joint project between the University of California Davis, Rothamsted Research, the Earlham Institute, and John Innes Centre. The populations consist of 1,535 Kronos and 1,203 Cadenza EMS-mutagenized M2 plants that have been exome-sequenced using Illumina next-generation sequencing. The sequences were mapped to the IWGSC RefSeq v1.0 assembly, and mutations were identified. The effects of the mutations can be predicted based on the RefSeqv1.0 protein annotation. As knock-out mutations are available in most wheat genes, the population provides an invaluable resource for identifying mutations in specific genes. Although these populations were originally produced with TILLING in mind and continue to be referred to as TILLING populations, the availability of nucleotide sequence data on the exomes of the entire populations means that TILLING for suitable mutations is no longer required.

5.1.4. Chapter aims

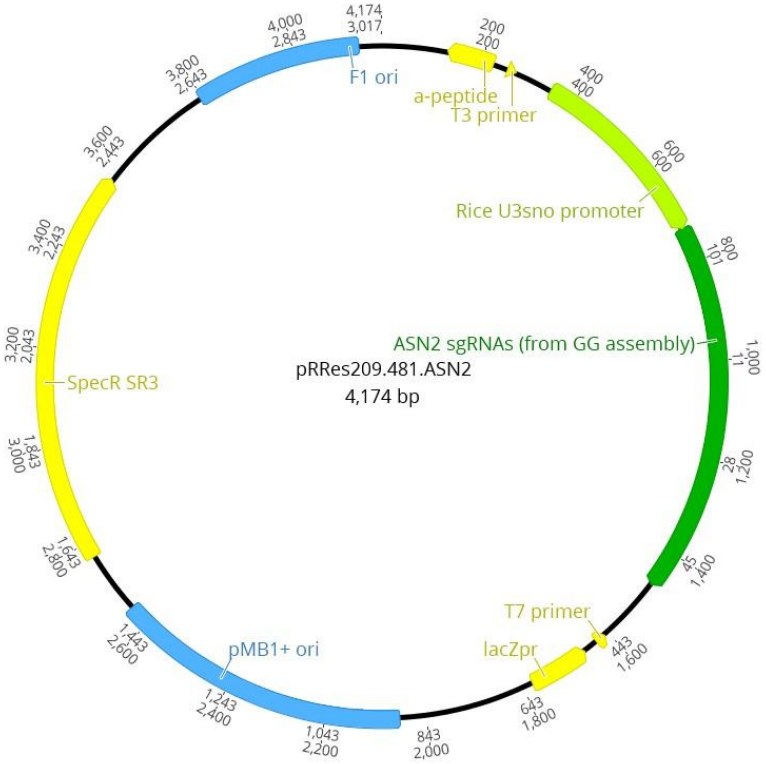
The aim of the work described in this chapter was to generate stably transformed genome-edited wheat plants with mutations in the *TaASN2* genes. In order to do this, isolated wheat cv. Cadenza embryos were subjected to biolistic transformation with the gRNA-containing plasmid, pRRes209.481.ASN2, alongside a *Cas9*-containing construct and a construct used for selection. T0 plants were regenerated and their DNA was isolated. PCR reactions were run to try to identify plants carrying large-scale editing events, such as large deletions or insertions.

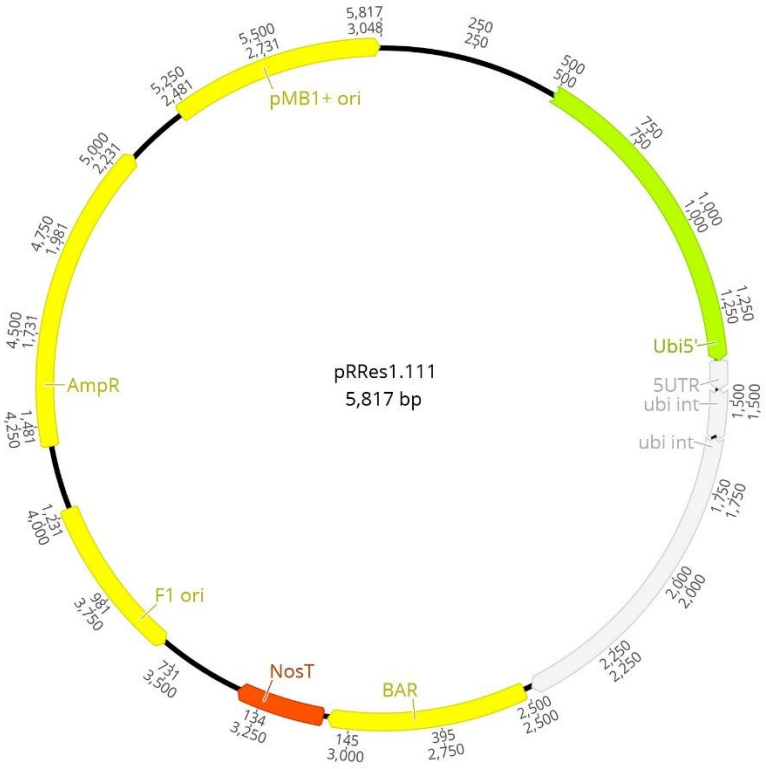
Alongside this, a wheat TILLING population was analysed for mutations in the *TaASN2* gene. Plants were identified carrying stop mutations in all six alleles (i.e. both alleles of all three genomes), with the aim of crossing these lines to generate a *TaASN2* triple null line.

5.2. Results

5.2.1. Transformation of *Triticum aestivum* cv. Cadenza embryos with the plasmid pRRes209.481

Immature embryos were transformed *via* biolistic bombardment. The transformation was conducted by Lucy Hyde at the Cereal Transformation Facility at Rothamsted Research. Three plasmids were co-bombarded into the embryos: pRRes209.482.ASN2; pRRes217.486, containing a Cas9-encoding construct; and the pRRes111.1 plasmid, which contains the *BAR* gene, encoding phosphinothricin N-acetyltransferase for tolerance of the herbicide Basta. Both additional plasmids (Figure 5.1) were obtained from Alison Huttly (Rothamsted Research).


pRRes209.481.ASN2
4,174 bp
F1 ori
a-peptide
T3 primer
Rice U3sno promoter
ASN2 sgRNAs (from GG assembly)
T7 primer
lacZpr
pMB1+ ori
SpecR SR3


pRRes1.111
5,817 bp
pMB1+ ori
Ubi5'
5UTR
ubi int
ubi int
AmpR
F1 ori
NosT
BAR

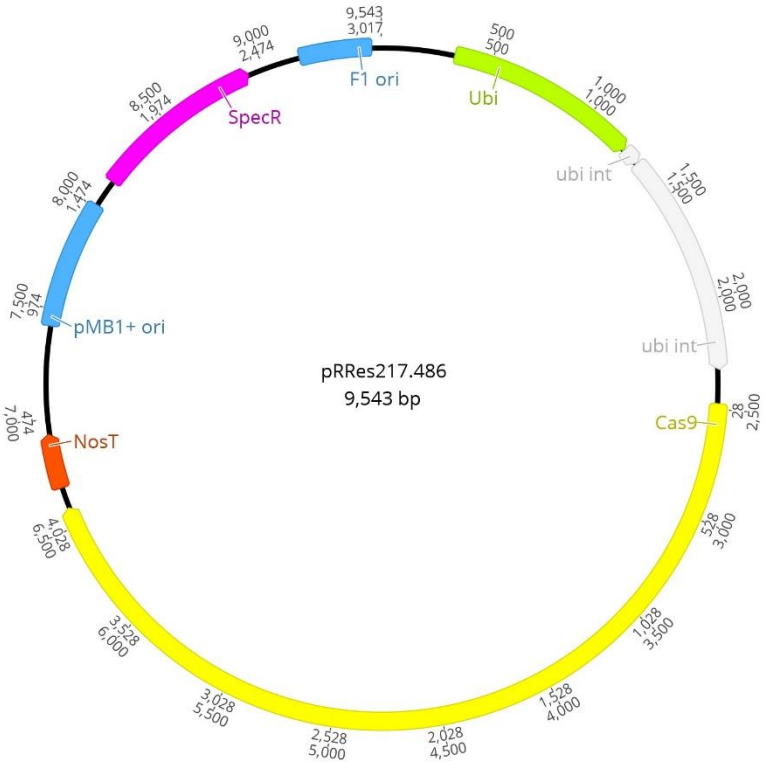


Figure 5.1. The three plasmids used in the project. Top: pRRes209.481.ASN2 (generated in Chapter 4). Middle: the pRRes111.1 plasmid, containing the selectable marker gene (*BAR*), obtained from Alison Huttly (Rothamsted Research). Bottom: pRRes217.486, with a construct containing the *Cas9* gene, obtained from Alison Huttly (Rothamsted Research).

A total of six bombardments were conducted, with the first two bombardments occurring on the same day. The following bombardments took place over many weeks and so only the plants from the first bombardment were taken forward. Plants from the following bombardments were planted out but discarded when the editing efficiency of the first two bombardments was assessed.

5.2.2. Analysing the T0 population for *Cas9* and *BAR* integration

The T0 plants were regenerated under selection and analysed for editing and transformation. The plants were initially genotyped for the presence or absence of the resistance marker and *Cas9* gene. The six bombardments generated a total of 92 plants, 77 (84 %) of which showed *Cas9* integration and 84 (91 %) *BAR* integration. Only eight plants showed *BAR* integration without *Cas9*, and only one plant showed *Cas9* integration without *BAR*.

5.2.3. Analysing the T0 population for editing events

Due to the lack of a convenient restriction site, restriction analysis to identify editing events could not be performed. Instead, plants were initially screened by amplicon amplification and band separation *via* gel electrophoresis, because large editing events would reduce the size of the amplicon. This method is limited in its ability to detect small edits, such as base substitutions, single (or small) base pair deletions or insertions and transversions. Generic primers (ASN2-2F and ASN2-1R, with an expected product size of 421-433 bp) were used to amplify the A, B and D genomes simultaneously, leading to six potential different editing patterns.

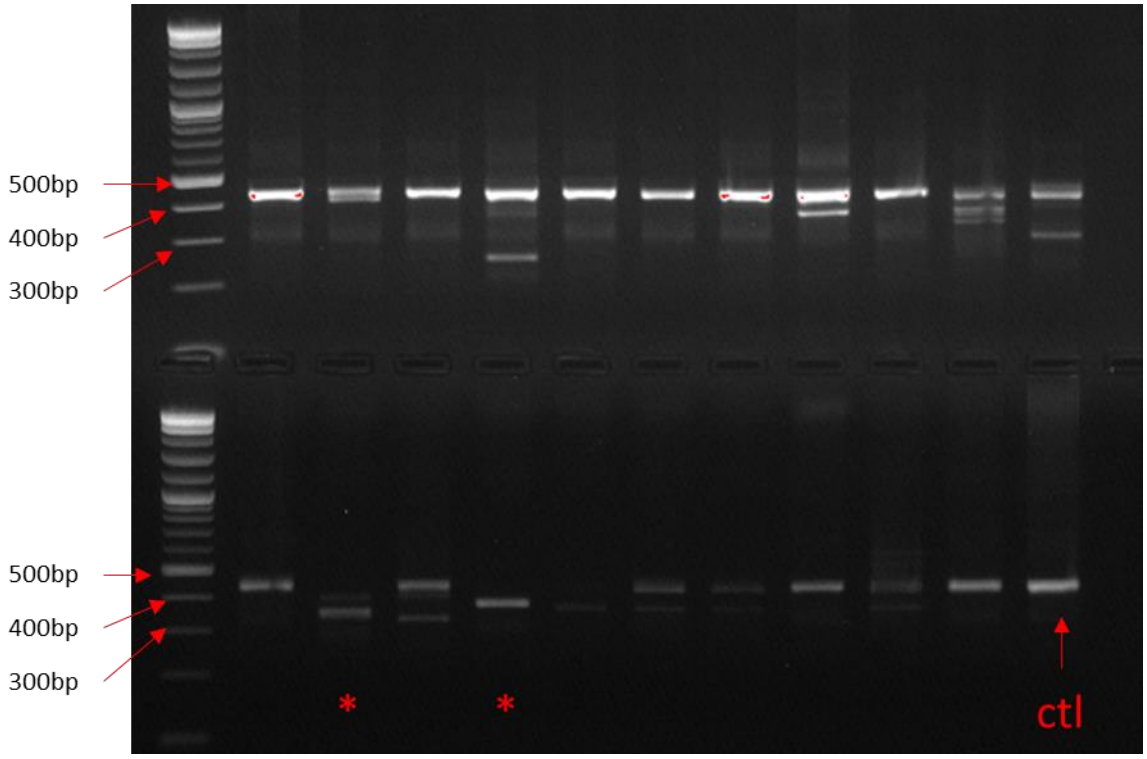


Figure 5.2. Electrophoresis gel showing the products of PCR amplification of the first exon of *TaASN2* in the T0 plants from the first bombardment, using the 2 × ReddyMix polymerase and the ASN2-2F and ASN2-1R primers. The expected band was around 430 bp and the PCR products were run against the 2-Log DNA Ladder (0.1-10.0 kbp) (New England Biolabs, Hitchin, UK). * denotes potential triple nulls.

From initial amplification of the target site, only two potential null mutants were identified in the T0 genome edited plants (Figure 5.2). However, A genome nulls were also of interest because RNA-seq data had shown the A genome copy of *TaASN2* to be much more highly expressed (> 3-fold difference) than the D genome copy (Section 3.3.2; Curtis *et al*., 2019). In fact, the A genome copy of *TaASN2* was responsible for more than half of the total asparagine synthetase gene expression (including *TaASN1-4*) in the grain under both sulphur sufficiency and deficiency and at both 14 and

21 days post-anthesis, suggesting that genetic interventions aimed just at this gene could bring about a substantial reduction in asparagine synthetase activity.

The samples were extracted from the gel and the two potential triple mutants were analysed for their nucleotide sequences. This identified a 62 bp deletion between the two inner gRNAs, although it was difficult to distinguish all three genomes from the sequence data and thus confirm the presence of mutations in all three genomes.

The PCR was re-run using a different polymerase (Q5) and forward primer (ASN2-F1, used with ASN-1R), showing a WT band in the potentially edited samples. This suggested that a visual band shift pattern was unlikely to be a good indicator for editing events (Figure 5.3). Similar PCRs were carried out for the remaining bombardments. Editing events in the T0 were not accurately assessed and consequently plants were taken forward from the first bombardment on the basis of survival and seed number for editing to be confirmed later.

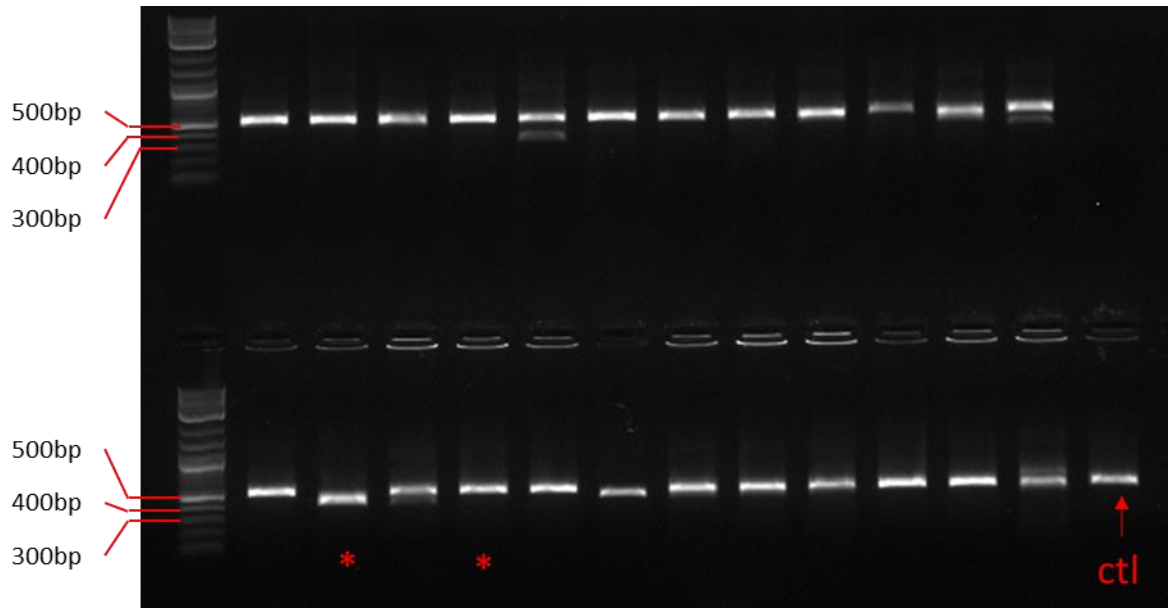


Figure 5.3. Electrophoresis gel showing secondary PCR amplification of the first exon of *TaASN2* in the T0 plants from the first bombardment, using the Q5 polymerase and the ASN2-F1 and ASN-1R primers. The expected band was around 430 bp and the 2-Log DNA Ladder (0.1-10.0 kbp) was used. * denotes the potential triple nulls.

5.2.4. Characterising the T0 edits through nucleotide sequence analysis of the T1 plants

Eventually the frequency of editing events in the T0 lineages was assessed from the T1 Next Generation sequencing (NGS) nucleotide sequence analysis (described in Chapter 6). Out of the 14 T0

lines that were taken forward, 11 of them showed some form of editing and 9 of the lines showed plants with edits in all three of the genomes (Table 5.1). A total number of 181 T1 plants were planted from the 14 T0 plants, with 9-20 plants sown for each line, depending on the available seed number. From this, 167 plants were subjected to NGS analysis (Chapter 6). Out of the original 14 T0 lineages, 11 showed editing present, with nine T0 lineages showing editing in all three genomes.

Table 5.1. The lineages and genotypes of the chosen T0 plants, and the T1 progeny that was planted from them. The number of T0 lines that were edited is assessed from the editing seen in the T1 plants. The number of T1 plants derived from the T0 lineages and the number subjected to NGS analysis is also shown. From this, the number of plants that show editing in any form, and the number of plants that show editing in all three homeoalleles simultaneously is recorded.

T0 Plant Lineage	T1 Plants	No. of T1 plants	No. of plants analysed	Editing present	Editing in all genomes
R3P1	12-23	12	11	11	11
R3P5	24-35	12	10	10	5
R5P3	36-47	12	11	9	0
R5P4	48-59	12	12	10	8
R5P6a	60-79	20	18	0	0
R6P1	80-91	12	11	0	0
R6P2	92-103	12	11	11	7
R9P1	104-123	20	20	18	2
R10P1	124-135	12	12	0	0
R14P2	136-147	12	9	8	6
R14P3	148-159	12	11	8	3
R15P1	160-171	12	11	11	7
R17P1	172-183	12	12	3	0
R17P3	184-195	9	8	8	6
	Total no.	181	167	107	55

5.2.5. TILLING mutants

Due to the regulatory issues regarding genome editing, and the commercial interest in the project, TILLING (Targeting Induced Local Lesions in Genomes) mutants were also investigated. Lines containing mutations resulting in the introduction of in-frame stop codons were identified in all three homeoalleles of the *TaASN2* gene in an EMS-mutated cv. Cadenza population (Krasileva *et al*., 2017).

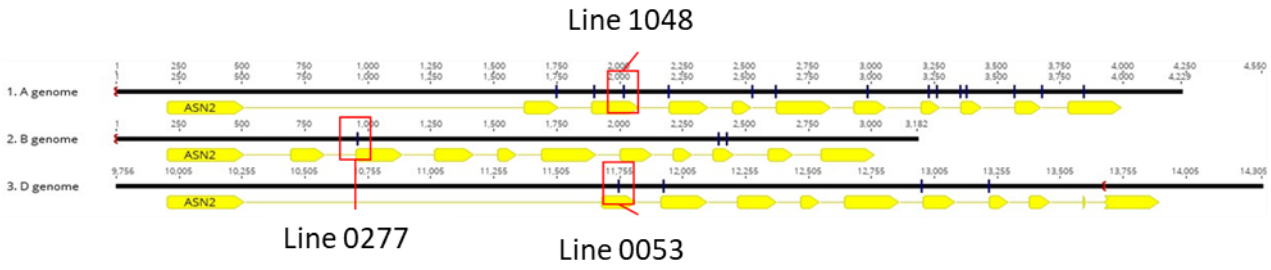


Figure 5.4. Diagram showing the mutations in mutant lines chosen from the cv. Cadenza TILLING population (Krasileva *et al*., 2017). Lines with mutations in the A, B and D homeologues are shown, and the exons of the *TaASN* gene are highlighted in yellow. The positions of the mutations introducing stop codons in the genes of the chosen lines are highlighted by a red box.

In the A genome, there was a single 'stop mutation' identified in the third exon of the *TaASN2* gene, but this mutation was shared by two lines: Line 1048, which was homozygous for the mutation, and Line 1002, which was heterozygous for the mutation. There were further mutations identified in the 8th (Line 1575), 9th (Line 1585) and 11th exons (Lines 1655 and 0469), which were all heterozygous. In the D genome, there were two stop mutations identified in the *TaASN2* gene. These were in Line 0053, which was homozygous for this mutation, and line 0764, which was heterozygous for the mutation. The mutation in Line 0053 was in the 2nd exon of the gene, whilst the mutation in Line 0764 was in the 3rd exon. A single stop mutation, in the 3rd exon of the gene, was identified for the B genome. This was in Line 0277, which was heterozygous for this mutation.

Accessions 1048, 0277 and 0053 were chosen to be introduced into a crossing programme, conducted at one of the project partners, RAGT, so that the mutations could be stacked to give triple and partial null mutants, with specific interest in the A genome null mutant. It was decided to stack the mutations in variety Claire, which, unlike Cadenza, is one of the varieties already lacking a B genome *TaASN2*, in order to reduce the number of crosses required. This work is ongoing, with an A/B genome null line expected to be ready by June 2021.

Mutations were also identified in the tetraploid cv. Kronos population. In the A genome, four lines were identified with stop codons in the *TaASN2* gene. These were Line 2032 (exon 2, heterozygous), Line 1388 (exon 3, heterozygous), Line 3562 (exon 7, heterozygous) and Line 3629 (exon 8, homozygous). In the B genome, two lines were identified, both carrying stop mutations in the 9th exon. These were Lines 2395 and 2646, and both lines were heterozygous for the mutation.

5.3. Discussion

5.3.1. Transformation of the wheat plants and analysis of the T0 generation

Edit detection remains difficult in genome edited plants (Grohmann *et al*., 2019; Zhang *et al*., 2019), with multiple methods often employed to characterise the mutations generated (Kim *et al*., 2018; Zhang *et al*., 2019). The initial analysis for edits in the T0 plants produced in this study relied on amplification of the target site, with the hope that successful editing would produce a size difference in the subsequent amplicon (Nekrasov *et al*., 2017). This skewed edit detection towards 'knock-outs' between the gRNAs; however, there are a variety of smaller edits that would lead to loss of function of the encoded protein, such as single base pair deletions or insertions, or non-synonymous substitutions. Furthermore, this method of edit detection did not appear to be particularly reliable, giving inconsistent results depending on the polymerase used. Thus, the edits were investigated using NGS to better understand the precise nature of the edits present in the T1 generation (Chapter 6). As editing could not be accurately assessed in the T0 generation, plants were taken forward to the T1 generation based on survival and seed number, rather than on the presence of successful editing events. Whilst the PCR reactions to amplify the first exon of *TaASN2* were carried out for the subsequent bombardments, the plants were not taken forward, due to the success of the original (first bombardment) plant lines. As the biolistic bombardments were carried out over a period, by the time that the PCR reactions were conducted to assess the last of the T0 plants, the T1 plants from the first bombardment were already being analysed by NGS and it was becoming clear that editing had been achieved (Chapter 6).

The presence of editing in the T0 lines was later estimated from the edits seen in their T1 progeny (Table 1); however, this does not account for the potential trans-generational CRISPR/Cas9 activity and so may not accurately reflect the presence of editing in the T0 lines. This showed an extremely high frequency of editing in the T0 generation, with 11 out of the 14 lines showing some form of editing present (78.6 %) and nine lines showing editing in all three homeologues (64 %). Five of the T0 lines (R3P1, R3P5, R6P1, R15P1, R17P3) showed editing in all of the T1 plants, and one line, R3P1, showed editing in all three genomes in 100 % of the T1 plants. This is higher than has previously been reported in wheat. Wang *et al*. (2014) obtained a mutational frequency of 5.6 % when targeting the A genome homeologue of the *TaMLO* locus. Okada *et al*. (2019) obtained biallelic mutations at an efficiency of only 5 % when targeting the *Ms1* gene in wheat. Zhang *et al*. (2016b) saw editing efficiencies of up to 10 % through CRISPR/Cas9 DNA or RNA transient expression, although lower frequencies are expected when selection is not used. The high mutational rate achieved in this study could be due to simultaneous targeting of one gene by 4 gRNAs alongside use of a wheat-optimised *Cas9* gene.

The biolistic bombardments of the three plasmids showed successful co-transformation of the wheat embryos, with 76 plants showing integration of both the *Cas9* gene and the *BAR* gene, which was expected as the plants were selected for BASTA resistance. The integration of the gRNA-containing plasmid has not been determined; however, given that 11 out of the 14 T0 lines showed successful editing, it is likely this also showed high levels of integration. The copy number of these genes is unknown and, despite the fact that regulations covering genome-edited plants classify them as GMOs, regardless of the presence or absence of a transgene, it may be important to pursue transgene-free lines.

5.3.2. Selecting TILLING mutants

The TILLING lines are of particular use to the project's plant breeder partners, given the current uncertain regulatory situation for genome edited crops in the UK and European Union. The genome editing focused on targeting mutations to the first exon of the *TaASN2* gene, which encodes components of the glutamine-binding domain of the enzyme, in an attempt to render the enzyme completely dysfunctional. Targeting is not possible with chemical mutagenesis, of course; however, TILLING lines were identified that possess stop mutations in the 2nd and 3rd exons. The proteins encoded by these mutant alleles might possess functioning glutamine-binding domains (Xu *et al*., 2018), but it is highly unlikely that they could contribute to asparagine biosynthesis since they would lack the aspartate-binding domain. Mutations were also identified in the tetraploid cv. Kronos population; however, these have not so far been used in a breeding programme. The Cadenza mutations have been crossed into the Claire background for ease of use with the industry partners. Claire already lacks a B genome version of *TaASN2*, like Chinese Spring (Xu *et al*., 2018), reducing the number of crosses required to generate a null genotype.

Chapter 6

Characterisation of Editing Events using Next Generation Sequencing (NGS) Nucleotide Sequence Analysis

6.1. Introduction

CRISPR/Cas9 editing can be of low efficiency, and this, coupled with the complexity of the wheat genome and the fact that not all edits are created functionally equal, meant that the presence of editing events in the transformed plants needed to be confirmed, and the edits assessed for their functional effect. A single base pair substitution in the third base of a codon, for example, may have no effect on the encoded protein due to the redundancy of the triplet code. Similarly, the deletion of a whole codon could also have little effect if the encoded amino acid were not important in protein function or folding. Enzymes TaASN1 and TaASN2, for example, only show around 85 % similarity in the nucleotide sequences encoding them, yet have similar biochemical properties (Raffan and Halford, 2021; Xu *et al*., 2018).

6.1.1. Edit analysis by nucleotide sequence analysis

As mentioned in Chapter 5, further characterisation of the edits present in the transformed plants was needed. More than the presence and absence, the functional effect of the edits needed to be characterised. The methods described in Chapter 5 could only detect the presence of some types of mutation and could not determine the nucleotide sequences of the mutated genes, and so did not give a detailed characterisation of the edits or their predicted effects. One method to achieve this is to clone and analyse the undigested product from the PCR-RE analysis (Section 4.2.5) (Feng *et al*., 2016). This method has already shown success in wheat when targeting the *TaMLO* genes (Wang *et al*., 2014), with the mutations being confirmed by nucleotide sequence analysis. However, this strategy restricts target sequences to those that contain a restriction enzyme site. Despite the use of the *Sex*A1 restriction site in the protoplast system (Section 4.2.5), this was not applicable to the edited wheat lines because gRNA2 was the only gRNA that overlay a restriction site.

An alternative strategy would be amplification of the target site *via* PCR with genome-specific primers, and then nucleotide sequence analysis of the purified amplicons from each individual plant *via* Sanger sequencing to provide information on the edits present. However, direct nucleotide sequence analysis of the PCR amplicons would result in superimposed sequence chromatograms that would be difficult to interpret. Another alternative would be the amplification of the target site *via* PCR, cloning of the amplified bands, and analysis of the individual cloned fragments though Sanger nucleotide sequence analysis (Howells *et al*., 2018). Had a suitable restriction site been present at the target site, the DNA could have been enriched using a restriction digest (Shan *et al*., 2013). Due to the number of plants to be analysed, the polyploid nature of wheat and the variety of potential edits that could be present, the cost and time to complete the analysis this way was determined to be too high and other methods were available. Ma *et al*. (2015; 2016) and Liu *et al*. (2015) reported a degenerate sequence decoding (DSD) method which was used to identify the mutated sequences from

superimposed chromatograms in rice and Arabidopsis, potentially bypassing the tedious and expensive cloning process. Wheat, however, still presents a particularly difficult challenge when it comes to edit identification due to its polyploid nature.

High-throughput nucleotide sequence analysis, such as Next Generation Sequencing (NGS), of the whole genome or single/multiple PCR amplicons allows for the detection of low frequency mutations, and has even been used to detect somatic mutations (Fauser *et al*., 2014; Feng *et al*., 2014). It is also useful for identifying possible off-target effects (Ueta *et al*., 2017). NGS nucleotide sequence analysis has been increasingly used for the analysis of CRISPR/Cas9 genome editing experiments. The method has been employed in a variety of plant species, including Arabidopsis (Fauser *et al*., 2014; Feng *et al*., 2014; Kim *et al*., 2016), a wild tobacco species (Kim *et al*., 2016), tomato (Ueta *et al*., 2017), rice (Lu and Zhu, 2017) and wheat (Wang *et al*., 2018). However, despite the decreasing costs of nucleotide sequence analysis, it is often seen as an expensive and complex method. Furthermore, it generates a large amount of data to be processed and requires specialist bioinformatic analysis. Multiplexing the NGS is a cost-effective way to screen multiple plants simultaneously through pooled barcoded amplicons. Barcoded DNA libraries are generated in one of two ways, either through ligating adaptor sequences to fragmented DNA, or by designing PCR primers that incorporate the barcode and adaptor sequences, so that the barcodes and the appropriate adaptors are added to the PCR product during the amplification process (Bell *et al*., 2014).

6.1.2. Chapter Aims

The aim of the work described in this chapter was to characterise the edits present in the T1 and T2 plant generations, with the objective of identifying edited lines to progress forward. Ideally, these lines would have been homozygous for an edit that was predicted to impair TaASN2 function, and would lack the *Cas9* gene due to it being segregated away. The detection of mutations was performed using Next Generation Sequencing (NGS) nucleotide sequence analysis of target site (first exon of the *TaASN2* gene) amplicons. The samples were bar-coded using adaptors of four bases between the target-specific primer and Illumina TruSeq adaptors (Table 1), allowing multiple samples to be pooled in the NGS analysis and reads assigned to the plants they were derived from using the barcodes. The aim of the NGS analysis was to identify T1 and then T2 plant lines to take forward for further analysis, including A genome null and triple null lines. The A genome nulls were desirable as it was unknown what effect the triple null mutations would have on plant function and health. The A genome gene was identified as being the most important gene for asparagine synthetase gene expression in the grain of wheat (Figure 3.2; Curtis *et al*., 2019) and it was proposed that knocking out the A genome gene alone may be enough for a step reduction in asparagine accumulation in the grain whilst minimising the risk of negative phenotypic effects (Luo *et al*., 2019).

6.2. Results

6.2.1. Genotyping the T1 plants via Next Generation nucleotide sequence analysis

167 potentially edited T1 plants were analysed by NGS, comprising between 10 and 20 plants grown from seed of each of the 14 selected T0 lines, alongside 10 wildtype plants. Editing events were detected in 107 plants (64 % of the non-control plants), deriving from 11 of the 14 T0 plants. Of these, 55 plants (33 % of the total plants; 51 % of the edited plants) were edited in all three genomes, at least heterozygously, and these were derived from 9 of the 14 selected T0 lines. Edits were not detected in 37 plants (22 %) and the remaining 21 plants showed an insufficient read number to be analysed.

6.2.1.1. Information on NGS for T1 generation

A total of 171,510 reads were generated in the T1 analysis for the 167 plants analysed, resulting in an average of 969 reads per plant; however, the plants were not equally represented in the data. Fewer than 200 mapped reads were seen in 56 plants, 31.6 % of the plants tested, with 23 showing fewer than 100 reads and 8 showing fewer than 50 reads. The plants from the lineages R5P5, R9P1, R14P2 and R14P3 showed the lowest number of reads (Figure 6.1), with the lowest of all being R14P3 with an average read number of 127 reads per plant. None of these lineages were taken forward to the T2 generation. Plant 71 (R5P6) showed only 17 reads mapping to it, with only one read mapping to the B genome, seven reads mapping to the A genome and nine reads mapping to the D genome. The R6P2 lineage, from which Plant 99 was chosen to be advanced to the T2 generation, showed the highest number of reads with an average of 2121 reads per plant. The individual plant with the highest number of reads in the analysis was plant 80 (R6P1), which had 4771 reads mapping to it.

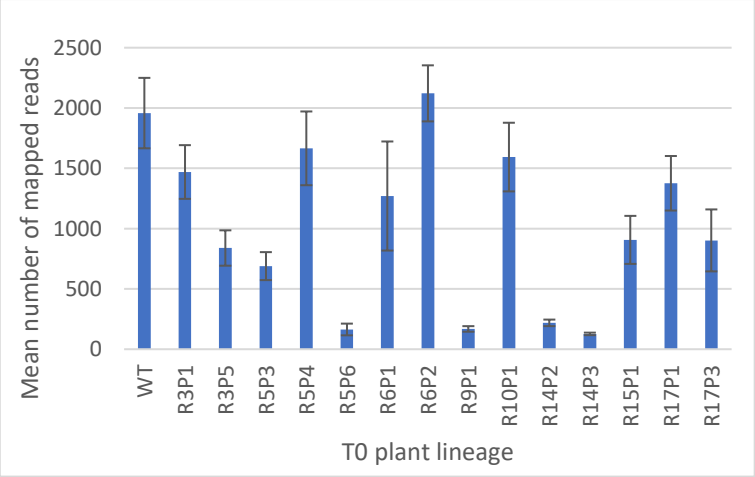


Figure 6.1. The mean read number mapping to each of the T0 plant lineages from the T1 NGS analysis. The error bars represent the standard error.

From this T1 NGS analysis, seven lines, 23, 30, 41, 59, 99, 126 and 178 were selected for further analysis (Table 6.1). The distribution of reads between the genomes was uneven (Table 6.1). The B genome was severely under-represented, accounting for 0.33 % of the reads in Plant 41, for example, with only two reads being detected. Out of the chosen T1 lines, it was most represented in Plant 126, where it accounted for 8.47 % of the reads. The A genome accounted for 46.82-64.70 % of the reads, with the D genome accounting for 34.72-50.50 % of the reads.

Table 6.1. The chosen T1 plants, with their T0 lineages, showing the total number of reads and the representation of the separate genomes, where the number of reads from each plant mapping to the different genomes is shown as a percentage of the plant's total read number.

T1 Plant	T0 Lineage	Total read no.	% of total reads in a plant		
			A	B	D
23	R3P1	737	53.9	5.2	41.0
30	R3P5	1688	62.4	2.9	34.7
41	R5P3	609	64.7	0.3	35.0
59	R5P4	3253	46.8	3.4	49.7
99	R6P2	3475	50.8	6.6	42.7
178	R17P1	501	48.3	1.2	50.5
126	R10P1	1831	48.8	8.5	42.7

6.2.1.2. Chosen Plants from the T1 NGS analysis

From these nucleotide sequence data, T2 seed from the T1 plants were chosen to be taken forward, forming the chosen plant lines (Table 6.2). Three putative 'triple' nulls were chosen: plant 23, plant 30 and plant 59. These plants showed editing in both alleles of all three genomes. The B genome status of plant 99 (*WT in Table 6.2) was undetermined due to the low read number. Two A genome nulls were chosen, plant 41 and plant 178, which showed editing in the A genome but no editing in the B or D genome. A wildtype-like plant (126) was chosen (not shown in Table 6.2) as another control, as this plant showed no editing events but had been subjected to the same transformation and growth procedures as the edited plants.

Table 6.**2.** Editing events in the *TaASN2* alleles and homeoalleles of the T1 generation of selected lines of wheat (*Triticum aestivum*) cv. Cadenza after editing with CRISPR/Cas9. The edits at the four gRNA positions, gRNA1-4, are shown for every genotype. Insertion and deletion events are represented by + and – signs, respectively, together with the number of additional or missing nucleotides. WT = wildtype. G → T indicates a guanine to thymine substitution. ND = not determined due to ambiguous data close to the end of the reads. * = provisional assignment based on poor coverage in the NGS data.

Line	Status	Genome	Zygosity	Allele	gRNA position			
					1	2	3	4
23	Triple null	A	Heterozygous	1			+1	+1
				2			G→T	+1
		B	Heterozygous	1				+1
				2			+1	+1
		D	Heterozygous	1			+1	-18
				2				-2
30	Triple null	A	Homozygous	1 & 2			+1	-1
		B	Homozygous	1 & 2			+1	-1
		D	Homozygous	1 & 2				-1
59	Triple null	A	Heterozygous	1	-11	+1	-2	ND
				2	+1	-4	+1	
		B	Heterozygous	1			G→T	ND
				2				+1
		D	Homozygous	1 & 2	-4	-11	+1	
99	A and D null	A	Heterozygous	1				+1
				2			-1	-2
		B	Homozygous	1&2	WT*			
		D	Heterozygous	1				+1
				2			+1	-1
41	A genome null	A	Homozygous	1 & 2	-14		+1	ND
		B	Homozygous	1 & 2	WT			
		D	Homozygous	1 & 2	WT			
178	A genome null	A	Homozygous	1 & 2			+1	
		B	Homozygous	1 & 2	WT			
		D	Homozygous	1 & 2	WT			

Plant 23 showed different edits in all six alleles (edits shown in Table 6.2). The first A genome allele showed the addition of a single base at both the gRNA3 and gRNA4 editing positions (cytosine and thymine respectively), introducing a stop codon in between the two gRNA positions in the nucleotide sequence. This allele encoded a truncated protein of only 81 amino acids, which was different to the wildtype protein from the 59th residue. One of the B genome alleles showed the same editing pattern. The second A genome allele showed a base change at the gRNA3 position, where a guanine had been substituted for a thymine in the forward strand, changing the encoded amino acid residue from a glycine to a valine. This amino acid is located outside the active site of the enzyme; however, glycine is known to be particularly important in protein folding and the addition of the hydrophobic sidechain of valine could also impact protein function (Betts and Russell, 2003). In any case, a single base (thymine) was also added at the gRNA4 target site, leading to premature termination of the amino acid sequence and the formation of a truncated protein of 97 residues.

The second B genome allele showed only the addition of a thymine at the gRNA4 position, resulting in the codon encoding the 98th residue being changed to a stop codon, again resulting in the formation of a truncated protein of 97 residues. The D genome also had two edited alleles, with one showing the loss of two bases at the gRNA4 position, resulting in a stop codon for the 97th codon and a truncated protein of 96 residues. The second allele showed the addition of an adenine at the gRNA3 position, resulting in a frameshift and the introduction of a stop codon between the gRNA3 and gRNA4 positions. There was a further loss of eighteen bases at the gRNA4 position. This resulted in a truncated predicted protein of 81 residues, with the first 59 amino acids remaining unchanged from the wildtype protein, followed by 22 changed residues.

Plant 30 was homozygous in each genome, with the A and B genome alleles showing the same edits. These comprised the addition of an adenine at the gRNA3 position, resulting in the disruption of the encoded protein from residue 59 onwards and a stop codon being introduced, producing a predicted protein of only 81 residues. There was an additional deletion of one base at the gRNA4 position, which would have returned the sequence back into frame if not for the presence of the stop codon. The D genome alleles showed the singular edit of the loss of a single base at the gRNA4 target position, causing a frameshift in the sequence, with translation ending at the next predicted in-frame stop codon at codon 161.

For Plant 59, both A genome alleles were shifted out of frame following the first gRNA target site, with the deletion of eleven bases in the first allele leading to a change in the predicted encoded amino acid sequence from the 11th residue. A single adenine was added at position gRNA2, leading to an overall loss of 10 bases. The sequence returned to the original reading frame at gRNA3, due to the addition of another adenine, resulting in the overall loss of three residues in the predicted amino acid sequence; however, a further two bases were deleted at gRNA4 and this resulted in a truncated

predicted protein of 93 residues. The second A genome allele showed the addition of a single thymine leading to the disruption of the encoded amino acid sequence from the 14th residue. The sequence reverted to the original reading frame with the deletion of four bases at the gRNA2 position, resulting in the loss of a single amino acid overall; however, the addition of an adenine at the gRNA3 position resulted in a frameshift and the introduction of a stop codon at the 81st codon.

The B genome possessed two alleles. In the first, a guanine had been substituted for a thymine and this resulted in a change from glycine to valine in the predicted amino acid sequence. The deletion of two bases at gRNA4 led to the introduction of a stop codon shortly afterwards (97th codon). The second allele possessed an additional adenine at the gRNA3 position, leading to disruption of the amino acid sequence and the introduction of a stop codon at codon 82. Only one D genome allele was present in the sequence data, suggesting that plant 59 was homozygous for this allele. The deletion of four bases at gRNA1 resulted in a frame shift in the nucleotide sequences, with the deletion of a further eleven bases at gRNA2 returning the sequence back to its original frame. The addition of a thymine at gRNA3 resulted in the introduction of a stop codon and, therefore, the production of a truncated predicted protein of only 76 residues.

Plant 99 showed two different edited alleles in the A and D genomes. In the first A genome allele, the addition of an adenine at the gRNA4 position led to the introduction of a stop codon at the 98th codon. This edit was also seen in the D genome, with the addition of an adenine at the gRNA4 position. The second A allele showed a single base (guanine) deleted at the gRNA3 position, leading to a frameshift; however, the deletion of two bases at gRNA4 returned the sequence to frame, resulting in a single amino acid lost overall and 37 residues changed in the predicted amino acid sequence, including known components of the glutamine binding domain (Gaufichon *et al*., 2010; Xu *et al*., 2018). The second D genome allele showed a thymine added at gRNA3, resulting in the addition of a stop codon at the 82nd codon, with an additional single base deleted at gRNA4. The B genome alleles were tentatively assigned to wildtype, although the NGS coverage of those alleles was poor.

Plant 41, one of the A genome null plants, showed wildtype sequences for the B and D genomes, although only two B genome reads were detected. Only one A genome allele was identifiable in the 4541 reads, indicating that plant 41 was homozygous for this edit. Fourteen bases were deleted at the gRNA1 editing position, leading to a frameshift and a change in the predicted amino acid sequence. Further edits were identified, with the addition of a thymine at gRNA3 and the loss of three bases at gRNA4; however, these edits did not return the sequence to the original reading frame. Plant 178, the second chosen A genome null plant, also showed wildtype sequences for the B and D genomes. Like plant 41, only one A genome allele was identifiable in the 5492 reads. This allele showed a single base introduction (an adenine) at gRNA3, leading to a stop codon at codon 82, resulting in a truncated protein of 81 residues that differed from the wildtype protein from the 59th residue. Plant 126 showed

no editing events in the target region and was taken forward as an additional control to the wildtype plants.

6.2.2. Genotyping the T2 plants *via* Next Generation nucleotide sequence analysis

6.2.2.1. Information on the T2 NGS analysis

A total of 5,744,714 reads were generated in the analysis, with an average value of 82,067.3 reads corresponding to each plant (Figure 6.2). A total of 1,833,600 reads were mapped to the A genome and 2,811,563 to the B genome, whilst only 1,099,551 mapped to the D genome. The number of reads per plant was a lot higher than in the T1 analysis; however, fewer plants were analysed (70 compared to 167) and the samples were not multiplexed in with other external samples.

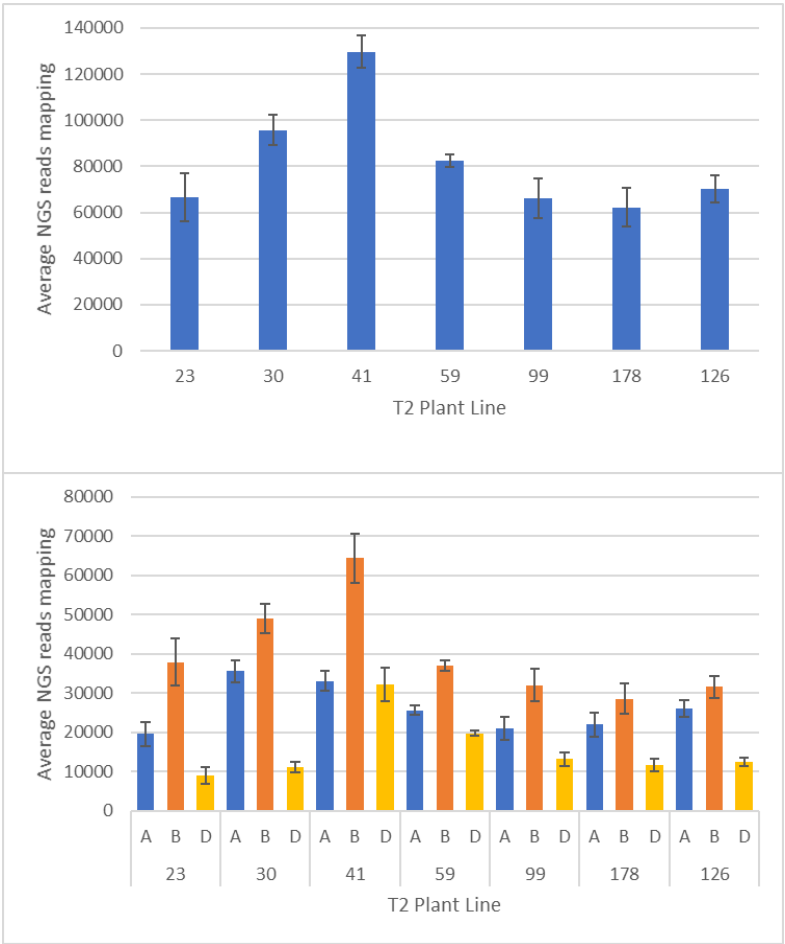


Figure 6.2. The number of NGS reads in the T2 analysis for each plant line mapping to the plant as a whole (top), and to the individual genomes in a plant (bottom). The bars represent the average mapped read number per line and the error bars represent the standard error.

Line 41 showed the highest number of mapped reads, with an average of 129,667.8 reads corresponding to each plant. The plant with the most reads in the analysis, Plant 82 (Line 41), showed 153,648 reads. The plant with the lowest number of reads was Plant 63 (Line 99), which showed only 2,426 reads: less than 2 % of the number of reads seen in Plant 82. The Line with the lowest number of reads was Line 178, with an average of 62,287 reads per plant. The difference in the mapped read numbers of individual plant samples may be due to a difference in the quality of the sample DNA. Line 41 shows a higher number of reads mapping than the other plant lines and this could be due to the large deletions present in this line resulting in shorter amplicons, which may amplify more efficiently.

6.2.2.2. NGS analysis of the T2 lines

The editing events observed in the T2 generation of lines 23, 30, 41, 59, 99 and 178 are summarised in Table 6.3, while the length in amino acids and molecular weights of the encoded proteins are given in Table 6.4.

Table 6.3. Editing events in the *TaASN2* alleles and homeoalleles of the T2 generation of selected lines of wheat (*Triticum aestivum*) cv. Cadenza after editing with CRISPR/Cas9. The edits at the four gRNA positions, gRNA1-4, are shown for every genotype. Insertion and deletion events are represented by + and – signs, respectively, together with the number of additional or missing nucleotides. WT = wildtype. G → T indicates a guanine to thymine substitution.

Line	Status	Genome	Zygosity	Allele	gRNA position			
					1	2	3	4
23	Triple mutant	A	Heterozygous	1			+1	+1
				2			G→T	+1
		B	Heterozygous	1		-173		
				2				+1
		D	Heterozygous	1			+1	
				2	WT			
30	Triple mutant	A	Homozygous	1 & 2			+1	-1
		B	Homozygous	1 & 2		-3		
		D	Homozygous	1 & 2				-1
41	Triple mutant	A	Heterozygous	1	-14		+1	
				2	-14	-27	+1	
				3	-14	-35	+1	

				4	-14	-59		
		B	Heterozygous	1	-75		-7	
				2	+1			-1
		D	Heterozygous	1	-74		+1	
				2		-60		
59	Triple mutant	A	Heterozygous	1	-11	+1	-2	
				2	1	-4	+1	
		B	Homozygous	1 & 2			G->T	-1
		D	Homozygous	1 & 2	-4	-11	+1	
99	AD null	A	Heterozygous	1				+1
				2			-1	
		B	Homozygous	1	WT			
		D	Heterozygous	1				+1
				2			+1	-1
178	A genome null	A	Homozygous	1 & 2			+1	
		B	Homozygous	1 & 2	WT			
		D	Homozygous	1 & 2	WT			

Table 6.4. The lengths and molecular weights of the predicted amino acid sequences from the nucleotide sequence analysis of the edited plants. Only edited alleles are shown for the plant lines.

Line	Genome	Zygosity	Allele	Length of predicted amino acid sequence	Predicted molecular weight (kDA)
WT	A	N/A	1 & 2	581	65.200
	B	N/A	1 & 2	581	64.982
	D	N/A	1 & 2	581	65.058
23	A	Heterozygous	1	81	9.015
			2	97	10.972
	B	Heterozygous	1	39	4.346
			2	160	17.737
	D	Heterozygous	1	81	9.162
			2	WT	
30	A	Homozygous	1 & 2	81	9.073
	B	Homozygous	1 & 2	580	64.899
	D	Homozygous	1 & 2	160	17.892
41	A	Heterozygous	1	156	16.515
			2	145	15.332

			3	557	62.086
	B	Heterozygous	1	133	14.128
			2	81	8.949
	D	Heterozygous	1	49	5.092
			2	561	63.083
59	A	Heterozygous	1	577	64.064
			2	80	8.611
	B	Homozygous	1 & 2	160	17.780
	D	Homozygous	1 & 2	76	7.960
99	A	Heterozygous	1	160	17.388
			2	97	10.930
	B	Homozygous	1	WT	
	D	Heterozygous	1	81	9.132
			2	97	10.946
178	A	Homozygous	1 & 2	82	9.073
	B	Homozygous	1 & 2	WT	
	D	Homozygous	1 & 2	WT	

6.2.2.2.1. Line 126 (WT-like)

In Line 126, ten plants were analysed. The total reads mapping to each plant ranged from 35127 to 93857, with 17.1 - 19.1 % of total reads mapping to the D genome, 33.4 - 39.9 % to the A genome and 41.0 - 48.1 % to the B genome. All plants were wildtype in the A (95.2-99.1 % of A genome reads), B (94.1-99.4 % of B genome reads) and D (97.7-99.8 % of D genome reads) genomes. The reads not mapping to the wildtype alleles represented 0.9 - 4.8 % of A genome reads, 0.6 - 5.9 % of B genome reads and 0.2 - 2.3 % of D genome reads. They did not map to any consensus and showed no pattern, with a high number of unexpected SNPs (whereby the SNP does not occur in an expected site and is present in only one of a few reads), or errors, present.

6.2.2.2.2. Line 23

In Line 23, ten plants were analysed (numbered 8, 12, 18, 33, 39, 45, 55, 60, 75 and 77). A huge range in the total number of mapped reads per plant was seen, from 6,650 for plant 39 to 110,098 for plant 60. The D genome showed the lowest number of mapped reads, with only 4.5 - 20.7 % of the total reads from a single plant mapping to the D genome. Interestingly, despite not being well-represented in the T1 generation NGS results, the B genome generally had the highest number of reads mapped to it, with 49.8 - 67.9 % of the total reads per plant compared with 20.7 - 38.0 % of the reads for the A

genome; although plant 12 showed 50.3 % of the reads mapping to the A genome, 44.8 % mapping to the B and 4.9 % (only 766 reads) mapping to the D genome.

Two distinct alleles were present in the A genome (edits shown in Table 6.3). The first A genome allele showed the addition of a single base at both the gRNA3 and gRNA4 positions (cytosine and thymine, respectively). This edit leads to the introduction of a stop codon in between the two gRNAs, leading to a truncated predicted protein of 80 amino acid residues (predicted proteins shown in Table 6.4), with a disruption of the amino acid sequence from the 59th residue. The second A genome allele showed a single thymine added at the gRNA4 editing site and the conversion of a glutamine to a thymine at position gRNA3. This edit resulted in the introduction of a stop codon after the gRNA4 position and a truncated predicted protein of 97 amino acid residues. Only two plants were homozygous for allele 1, plant 12 (97.6 % of A genome reads) and plant 60 (91.5 % of A genome reads), with the rest being heterozygous. Both of these alleles were seen in the previous generation.

Two edited alleles were also present in the B genome. The first allele showed a deletion of 173 bases between the gRNA2 and gRNA4 positions, and the second the deletion of a single base (cytosine) at gRNA4. The edit in the first allele would result in the loss of 57 amino acids in the predicted protein sequence, including key residues in the glutamine binding domain, and would result in a frameshift and the introduction of a stop codon resulting in a truncated predicted protein of only 39 residues. The second allele also showed a frameshift, disrupting the amino acid sequence in the encoded protein and leading to a truncated predicted protein of 160 amino acid residues. Again, plants 12 and 60 were the only two plants that were homozygous for allele 1 (88.3 % and 97.0 % of the B genome reads, respectively). The rest of the plants were heterozygous for both edited alleles, although 42.2 - 78.6 % of the reads mapped to allele 1 and only 17.9 - 55.5 % to allele 2. This may be a product of the size difference between the alleles as the amplicon from allele 1 is 173 bases shorter than that of allele 2 and thus could possibly amplify more efficiently. Neither of these edits were seen in the previous generation; however, only 45 reads mapped to the B genome in the T1 analysis, with only 737 reads mapping to plant 23 in total. Therefore, it is possible that the edits identified in the T2 generation were missed in the T1 analysis.

A single, novel, edited D genome allele was also present, along with wildtype reads. The edited allele had an additional adenine at the gRNA3 position, leading to an in-frame stop codon and an encoded protein of only 80 amino acids. Three plants (plant 8, 12 and 55) appeared to be homozygous for the wildtype allele, although plants 8 and 55 showed only 58.9 and 52.4 % of reads mapping, while plant 12 only showed a total of 766 reads mapping to the D genome (644 of which were wildtype). Plants 33 (92.9 % of D genome reads), 45 (89.8 %), 60 (89.6 %) and 75 (91.8 %) appeared homozygous for the edited allele. Plant 77 also only showed the edited allele; however, only 69.3 % of D genome reads mapped to it. Plants 18 and 39 showed a mix of wildtype and edited reads. Plant 18 only showed

15.0 % of D genome reads mapping to the wildtype allele, with 68.9 % of them mapping to the edited allele. Plant 39 showed 23.9 % of D genome reads as wildtype and 57.5 % reads as edited, although a total of only 923 reads were mapped to the D genome. The previous generation showed two edited alleles, neither of which were seen in the T2 analysis, although, again, this may be explained by the low mapped read number in the T1 analysis.

6.2.2.2.3. Line 30

In Line 30, there were ten plants tested (5, 11, 17, 27, 44, 48, 56, 73, 81 and 88). The plants showed a total mapped read number between 53,027 and 117,361. All 10 plants were homozygously edited in the A, B and D genomes, confirming the T1 plant 30 as a triple mutant. In the T2 plants, 96.6 - 98.6 % of the A genome reads, 89.9 - 97.5 % of the B genome reads and 89.2 - 99 % of the D genome reads mapped to the edited allele. Again, the D genome showed the lowest number of mapped reads, with only 8.1 - 18.7 % of the total reads from a single plant mapping to the D genome. The B genome had the highest number of reads mapped to it, with 45.7 - 56.0 % of the total reads mapping to it, with the A genome having 33.4 - 40.6 % of the reads.

The A genome showed the addition of a single base (adenine) at position gRNA3 and the deletion of a single base (cytosine) at the gRNA4 position. Although these edits do not change the total number of bases, they do result in a section of the first exon being 'out-of-frame' and the introduction of an in-frame stop codon (82nd codon).

The B genome allele showed the deletion of 3 bases at position gRNA2. Although this is an edit, it will not result in a frameshift, instead resulting in the deletion of a leucine in the translated sequence. As this particular residue has not been implicated as being important in protein function, it is unknown as to whether this edit would affect the activity of the encoded enzyme.

The D genome showed the deletion of a single base (cytosine) at the gRNA4 position. The predicted effect of this edit is a frame shift in the sequence, leading to stop codons being introduced after 161 codons and a completely different amino acid sequence being encoded after the 95th codon.

The A and D genome edits were consistent with what was seen in the T1 plants; however, the B genome allele was not detected in the previous generation. The B genome allele present in the T1 generation showed an identical editing pattern to that seen in the A genome, as opposed to the deletion of 3 bases at gRNA2, which was detected in this analysis. It is possible that the edited allele identified in the T2 analysis was missed due to the low read number of B genome reads in the T1 analysis (only 1688 reads in total, in which the B genome was substantially under-represented). The appearance of this and other edits in the T2 generation that were not seen in the T1 generation could also be evidence of continued Cas9 activity.

6.2.2.2.4. Line 41

In Line 41, ten plants were analysed (10, 15, 21, 29, 38, 43, 46, 61, 69 and 82). A total mapped read number of 90,958-153,648 was recorded per plant. The highest number of reads generally mapped to the B genome (37.6 - 62.2 % of total reads), with the exception of plant 29, where the D genome showed 41.7 % of the total mapped reads (27.8 % A and 30.5 % B genome).

One main allele was recorded in the A genome: the deletion of 14 bases at position gRNA1 and the addition of a single thymine at gRNA3, leading to disruption of the amino acid sequence in the predicted protein from the 14th residue. Seven of the plants were homozygous for this allele: plants 15 (94.1 % of the A genome reads for this plant mapped to this allele), 21 (93.5 %), 29 (95 %), 43 (99.9 %), 46 (96.5 %), 61 (96.6 %) and 69 (95.2 %). Plant 10 showed a second allele, with a deletion of 14 bases at position gRNA1, followed by the deletion of 27 bases at gRNA2 and the addition of a single thymine at gRNA3. However, this was only present in 16.1 % of the mapped A genome reads. Plant 82 also showed another allele, accounting for 63.9 % of the mapped A genome reads, where the deletion of fourteen bases at position gRNA1 was followed by the deletion of 35 bases at gRNA2 and the addition of a single thymine at gRNA3. This edit also leads to the disruption of the amino acid sequence in the predicted protein from the 14th residue. Plant 38 showed the same deletion of fourteen bases at position gRNA1 followed by a second deletion of 59 bases between the gRNA2 and gRNA3 positions, which was present in 58.4 % of the reads. This edit resulted in the same disruption from the 14th residue in the predicted protein sequence, with the amino acid sequence returning to the wildtype sequence after the 35th residue, with the total loss of 24 residues, including those important in the glutamine binding domain. All these alleles could be construed as evidence for continued editing.

Plant 41 was previously thought to be an A genome null only, with wildtype reads identified for the B and D genomes. Only one A genome allele was identifiable in the 4541 Next Generation reads in the T1 analysis, indicating that plant 41 was homozygous for this edit, which showed fourteen bases deleted at gRNA1, the addition of a thymine at gRNA3 and the loss of three bases at gRNA4. The loss of the three bases at gRNA4 was not seen in the T1 analysis, but the loss of fourteen bases at gRNA1 and the addition of a single base at gRNA3 was as seen in the T1 analysis. The presence of these four different A genome alleles is strong evidence of continued editing beyond the T0 generation. Furthermore, while the 14 base pair deletion at gRNA1 was present in the single A genome allele seen in the T1 generation, the longer deletions at gRNA2 were not, suggesting that these deletions were later edits added to the allele seen in the T1 generation. In addition, Line 41 was originally shown to be an A genome null, but by the T2 generation two edited alleles had appeared in each of the B and D genomes.

The B genome alleles showed no editing in the T1 analysis; however, the wildtype status of the B genome was not completely confirmed in the T1 analysis due to a low read number (609 mapped

reads total). Despite this, it was a surprise that two edited B genome alleles were identified in the T2 analysis: 1) showed a deletion of 75 bases between positions gRNA1 and gRNA2, followed by the deletion of 7 bases at gRNA3; and 2) showed the addition of a single base at gRNA1 (thymine) followed by the deletion of a single base (cytosine) at gRNA4. The first allele leads to a frameshift and the disruption of the amino acid sequence from the 14th residue. The second allele leads to the introduction of a stop codon and the truncation of the predicted amino acid after the 81st residue. Three plants were homozygous for allele 1: plants 10 (98.1 % of B genome reads mapped to the edited allele), 43 (97.8 %) and 61 (98.0 %). Two plants were homozygous for allele 2: plants 29 (94.8 % of B genome reads mapped to the edited allele) and 69 (97.5 %). The remaining plants were heterozygous for both edited alleles, although allele 1 was present in a lot more of the reads (77.6-82.5 % of B genome reads mapping to allele 1 compared with 16.2-21.9 % for allele 2). This could also be a product of the difference in amplicon length, as discussed in Line 23.

Two edited alleles were also identified in the D genome: 1) the deletion of 74 bases between positions gRNA1 and gRNA2, followed by the addition of a cytosine at gRNA3; and 2) the deletion of 60 bases between positions gRNA2 and gRNA3. The edits in the first allele result in the disruption of the amino acid sequence in the predicted protein from the 15th residue onwards and then the introduction of a stop codon, resulting in the amino acid sequence being truncated to just 49 residues. The deletion in the second allele results in the loss of 20 residues from the predicted protein sequence, including those comprising the glutamine binding domain. Four plants were homozygous for allele 1: plants 38 (94.6 % of D genome reads mapped to the edited allele), 43 (97.2 %), 46 (97.2 %) and 61 (98.7 %). Three plants were homozygous for allele 2: plants 10 (98.3 % of D genome reads mapped to the edited allele), 21 (99.5 %) and 29 (99.9 %). Three plants were heterozygous with both edited alleles: plants 15, 69 and 82, although all plants showed a higher number of reads mapping to allele 2 (58.8-66.5 % of D genome reads mapping to allele 2 compared with 31.1-39.0 % for allele 1).

6.2.2.2.5. Line 59

In Line 59, eleven plants were analysed (2, 6, 16, 26, 37, 54, 59, 67, 80, 84 and 89) and a total mapped read number of 66,885 to 95,531 was recorded for each plant. The highest number of reads mapped to the B genome (40.2 - 52.1 % of total mapped reads) and the lowest number mapped to the D genome (21.0 - 24.8 % of the total reads).

Two edited alleles were present in the A genome: 1) a deletion of 11 bases at gRNA1, followed by the addition of a single adenine at gRNA2, and the deletion of two bases at gRNA3; and 2) the addition of a thymine at gRNA1, the deletion of four bases at gRNA2 and the addition of an adenine at gRNA3. The first allele leads to a disruption in the amino acid sequence in the encoded protein from the 11th residue until the 56th residue, with the loss of four residues in total. The second allele leads to a disruption in the amino acid sequence in the predicted protein from the 14th residue until the

38th residue, with the loss of a single residue, and then again from the 58th residue until the introduction of a stop codon, resulting in a truncated predicted protein of 80 residues. Homozygous plants were only seen for allele 1; specifically, plants 16 (94.4 % of the A genome reads mapped to the edited allele), 26 (99.2 %) and 37 (97.6 %). The remaining plants were heterozygous, carrying both edited alleles (46.1 - 57.6 % for allele 1 and 41.5 - 47.8 % for allele 2).

The plants were all homozygous for a single edited allele in the B genome, with 94.4 - 99.6 % of the B genome reads mapping to this allele. A guanine was substituted with a thymine in the forward strand at position gRNA3 and a single base (cytosine) was deleted at the gRNA4 position. The latter leads to a frameshift and a change in the predicted amino acid sequence from the 96th residue onwards. Two alleles had been identified in the T1 generation and the edited allele present in the T2 generation did not match either of these exactly; however, it was very similar to the T1 allele 1, with a difference only in the number of bases deleted at the gRNA4 position. Due to the placement of the reverse primer used in the target site amplification being so close to the gRNA4 targeting site, it is possible that edits at the gRNA4 position were missed or incorrectly represented.

Plants were homozygous in the D genome for an edited allele with a deletion of four bases at gRNA1 and 11 bases at gRNA2, followed by the addition of a single thymine at position gRNA3 and the deletion of a cytosine at gRNA4, with 96.9 - 99.8 % of D genome reads mapping to this allele. This edit led to a disruption in the predicted protein sequence from the 13th residue onwards and the production of a truncated predicted protein of 76 residues. This matched the edited allele seen in the T1 generation, which was believed to be homozygous.

6.2.2.2.6. Line 99

In Line 99, nine plants were analysed (20, 24, 28, 50, 52, 58, 63, 68 and 72). The majority of the plants showed a total mapped read number of around 70,000 to 80,000 reads, with the exception of plant 50, which showed a total read number of 48,627, and plant 63, which showed only 2,426 mapped reads. Again, the D genome showed the lowest number of mapped reads, with only 17.9 - 22.7 % of the total reads from a single plant mapping to the D genome. The B genome had the highest number of reads mapped to it, with 45.6 - 51.3 % of the total reads, with the A genome having 26.8 - 35.4 % of the reads.

There were two different edited alleles present in the A genome: 1) the loss of a single base at gRNA3 (guanine), and three changed bases (CGG to GCA) at the gRNA4 editing position; and 2) the addition of an adenine at gRNA4. The edit in allele one leads to a frameshift and a disruption in the amino acid sequence from the 59th residue onwards. The edit in allele 2 leads to the introduction of a stop codon and a truncated predicted protein sequence of 97 residues. Only two plants were heterozygous for both edited alleles. Plant 20 showed 36.4 % of the mapped A genome reads corresponding to allele 1

and 62.0 % of reads mapping to allele 2. Plant 52 showed 42.2 % of reads corresponding to allele 1 and 57.4 % of reads mapping to allele 2. Plants 28, 50 and 58 were all homozygous for allele 1, with 92.2, 89.9 and 92.4 % of A genome reads mapping to that allele. Plants 68 and 72 were homozygous for allele 2, with 93.0 and 94.6 %, respectively, of their A genome reads mapping to allele 2. Plant 63 was also considered to be homozygous for allele 2; however, only 54.5 % of the A genome reads mapped to this allele. The first allele could not be detected and 8.1 % of the reads mapped to a different, potentially new, allele. These reads showed the addition of a cytosine at position gRNA3, and the addition of a thymine at position gRNA4. This could be an error due to primer position or could reflect continued editing and somatic mutations present in the plant. Plant 24 appeared to be heterozygous for allele 1 (38.3 % of mapped A genome reads) and wildtype (60.0 %), despite the fact that the wildtype allele was not seen in the A genome of any of the other Line 99 plants. This wildtype allele was not present in the T1 generation either. Due to the position of the reverse primer, which is very close to the gRNA4 position (shown in Figure 5), it is difficult to determine whether the deletion of the two bases seen in the T1 exists in the T2 generation but was not detected, or the edit did not originally exist, was somatic in origin or had segregated away.

All the T2 plants were wildtype in the B genome (91.5 - 98.8 % of B genome reads mapped to the wildtype sequence). In the D genome, there were two different edited alleles present: 1) the addition of a single base at gRNA3 (thymine) and the loss of a base (cytosine) at gRNA4; and 2) the addition of a single base (adenine) at gRNA4. Both these edits were present in the T1 generation. The edit in allele 1 leads to the introduction of a stop codon, and results in a truncated predicted protein of 81 residues. Allele 2 leads to the introduction of a stop codon and a predicted protein of 97 residues. Only two plants were heterozygous with both alleles. Plant 24 showed 53.6 % of the D genome reads corresponding to allele 1 and 44.3 % of reads mapping to the second allele. Plant 72 showed 51.6 % of reads corresponding to allele 1 and 46.5 % of reads mapping to the second allele. Plants 50, 58 and 68 were all homozygous for allele 1, with 95.8, 96.6 and 94.9 % of D genome reads mapping to the allele, respectively. Plants 28 and 52 were homozygous for allele 2, with 93.4 and 92.8 %, respectively, of reads mapping to allele 2. Plant 63 was also considered to be homozygous for allele 2, as this was the only detectable allele present; however only 37.6 % of the D genome reads mapped to this allele, with the rest of the reads not mapping to any consensus. Again, one of the Line 99 plants appeared to be heterozygous for allele 1 (52 % of mapped D genome reads) and wildtype (46.5 %), despite the fact that the wildtype allele was not seen in any of the other plants. This wildtype allele is not present in the T1 generation either, with both T2 editing alleles present in the T1 generation.

6.2.2.2.7. Line 178

In Line 178 there were 10 plants tested (3, 13, 23, 35, 42, 51, 64, 66, 76 and 86). Overall, the D genome showed the lowest number of mapped reads, with only 16.6 - 21.4 % of the total reads from a

single plant mapping to the D genome. The B genome had the highest number of reads mapped to it, with 38.3 - 50.6 % of the total reads mapping to it, with the A genome having 29.5 - 42.9 % of the total reads. All 10 plants were wildtype (homozygous) in the B (93.0 - 99.2 % of the reads were wildtype) and D (93.9 - 99.6 % of reads were wildtype) genome.

In the A genome, only one edited allele was observed, with the addition of a single base (adenine on the forward strand) at gRNA3, in 93.9 - 98.9 % of mapped A genome reads. This edit leads to a frameshift in the sequence and the introduction of a stop codon (82nd codon), which would result in a truncated protein being produced. This edit is consistent with the results seen in the T1 generation, although the wildtype status of the B genome was unconfirmed in the T1 data due to the low read number.

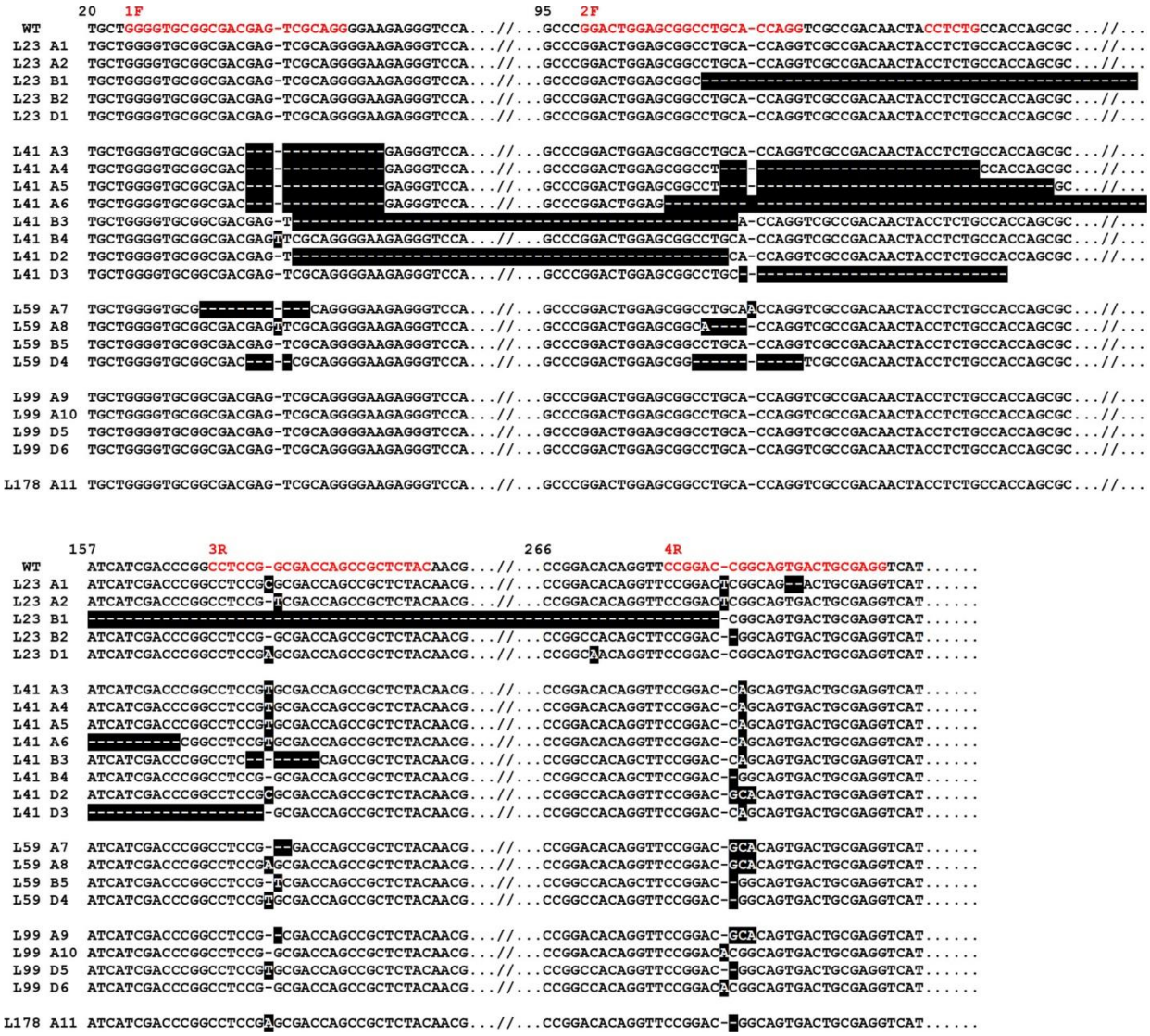


Figure 6.3. Nucleotide sequences showing the edited *TaASN2* alleles present in the T2 generation of wheat (*Triticum aestivum*) cv. Cadenza plants after editing with CRISPR/Cas9. Line 30 is not shown. The A genome wildtype sequence is shown at the top, with the gRNA binding sites in red. Edits are

highlighted in black. SNPs present in the wildtype sequences of the genes from genomes B and D are not highlighted. The different alleles are numbered, A1 to A11, B1 to B5 and D1 to D6.

6.3. Discussion

6.3.1. Comparison of the T1 and T2 analyses

There was a remarkable difference seen between the NGS analysis of the T1 and T2 generations. Firstly, there was a large difference in the read number of the two analyses, with the T1 analysis generating 171,510 reads in total, and the T2 analysis generating a total of 5,744,714 reads. The average read number per plant was 969 in the T1 generation compared with 82,067.3 for the T2 plants, which is nearly a 100-fold increase in coverage. The T1 analysis also showed a skewed distribution in read number, with a lot of plants showing a very low number of reads, whereas the T2 analysis showed the expected normal distribution in read number per plant (Figure 6.3).

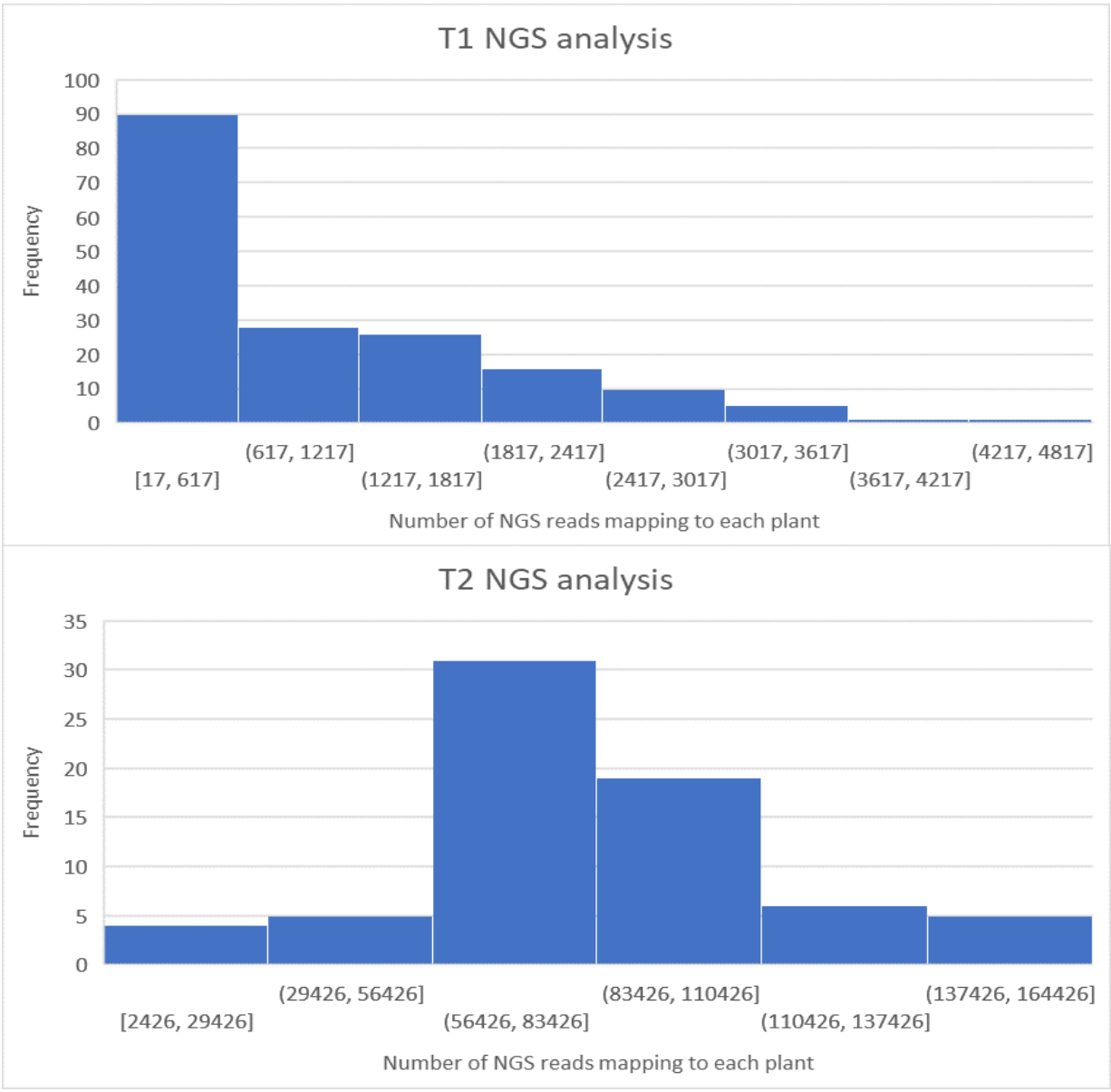


Figure 6.4. Histograms showing the number of reads mapping to each plant in the T1 (top) and T2 (bottom) Next Generation Sequencing nucleotide sequence analyses, highlighting the difference in sequence coverage between the analyses.

In the T1 analysis, the B genome was under-represented, with only two B genome reads present in Plant 41, for example, and only 0.3 - 8.47 % of the total reads in the chosen T1 plants mapping to the B genome. The A genome was generally the most highly represented genome in the analysis, with the D genome only being the most highly represented in two of the seven chosen lines. This was not the same in the T2 analysis, where the B genome was the most highly represented genome, accounting for 30.5 - 68.0 % of the reads per plant, with the A genome making up 19.0 - 50.3 % of the reads and the D genome 4.5 - 41.7 % of the reads. The same primers were used in both analyses, with the Next Generation sequencing being performed at the Genomics Facility at the University of Bristol.

6.3.2. Evidence for continued editing beyond the T0 generation

Only one edited line showed the same edits in the T2 and T1 generations. This was Line 178, which retained its initial status as an A genome null. A possible explanation for some of the differences in edits seen between the T1 and T2 generations is the difference in the read number per plant in the analyses. The increased number of reads per plant, therefore the greater coverage in the sequencing data, in the T2 analysis allowed for greater determination of the edits present. The low number of reads per plant in the T1 generation may have masked or misrepresented the edits present, particularly in the B genome.

Many more errors or unexpected SNPs (i.e. SNPs outside the editing sites and only identified in one or a few reads) were present in the T1 analysis, perhaps because of the lower read number per plant, making it harder to distinguish the exact number of reads mapping to each genome. Furthermore, the more heavily edited an allele is, particularly if the edits include large deletions, such as those seen in Lines 23 or 41, the harder it is to distinguish the A, B and D genomes in the reads, because the SNPs used to distinguish between the homeoalleles may have been deleted. It may be necessary in further identification to design new primers, or to amplify the A, B and D genomes separately.

Furthermore, the primers used in the amplification of the target site corresponded to positions that were close to the editing sites, in particular the reverse primer binding site was close to the gRNA4 target position (Figure 6.4). This may mean that some edits at the gRNA4 editing site were missed in the NGS analysis because of their proximity to the primer binding site. Options for primer location was limited due to deletions in the intron following the first exon of *TaASN2* (Chapter 4; Figure 4.11).

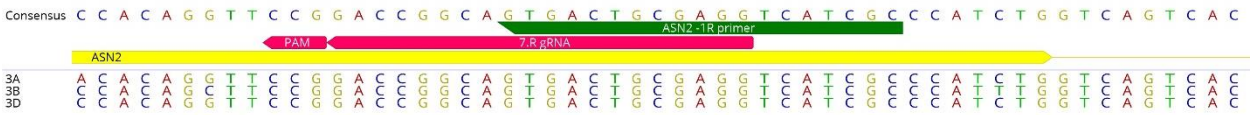


Figure 6.5. The position of gRNA4 and the reverse primer used in the PCR amplification of the target site. The two are highlighted on the consensus sequence generated from the A, B and D genomes, which are shown below it. The first exon is highlighted by the yellow annotation, showing the start of the first intron shortly after the primer site.

Not all of the changes in edits between the T2 and T1 generations could be explained by differences in the NGS coverage. Line 23, for example, showed wildtype alleles in the D genome as well as very different edits in the B genome to what was seen in the T1 data. One of the edits seen in the B genome was a 173 bp deletion between the gRNA2 and gRNA4 edition positions, which was not identified in the T1 analysis and was unlikely to have been overlooked. Line 30 remained a triple mutant in the T2

generation, as it was in the T1 generation, with the plant being homozygously edited in all three genomes. However, the B genome edit in the T2 generation was different to that seen in the T1 analysis. Line 59 also remained a triple mutant; however, differences in the specific edits were seen. The T1 plant was heterozygously edited in the B genome, whereas only one allele was seen in the T2 analysis, with differences seen in the edits at the gRNA4 position. Line 99, although designated WT for the B genome in the T1 analysis, did show some edited alleles, and so the B genome status was only assigned tentatively; however, it was clearly and unambiguously wildtype in the B genome in the T2 analysis, with all nine plants being homozygous for the wildtype allele. Wildtype alleles were also seen in the D and A genomes, respectively, of plants 20 and 24 of this line. There was also a difference in one of the A genome edited alleles, where a two-base deletion at the gRNA4 position seen in the T1 generation was not present in the T2 plants. Line 41 was also characterised as an A genome mutant in the T1 generation, with a homozygous edit seen in the A genome, a wildtype D genome and at least tentatively a wildtype B genome, although this was based upon only two single reads. In the T2 analysis, however, the Line 41 plants were all triple mutants, with large deletions of up to 75 bases in the B and D genomes. There was a difference seen between the T1 A-genome allele and the T2 as the deletion of three bases at gRNA4 seen in the T1 plant was not seen in the T2 progeny. As the three base deletion is not seen in any T2 plant, it is likely this was just an incorrect report of the T1 edited allele, as only 394 reads mapped to the A genome of T1 Plant 41, particularly with the proximity of the gRNA4 editing site to the reverse primer.

The discovery of novel genome editing events in the T2 progeny suggest ongoing and active CRISPR/Cas9 activity, a phenomenon that has been debated in previous wheat CRISPR/Cas9 experiments (Howells *et al*., 2018; Wang *et al*., 2018; Zhang *et al*., 2019). The transgenerational activity of the CRISPR/Cas9 system has been shown to induce new functional variants at multiple target sites in tomato (Rodríguez-Leal *et al*., 2017) and barley (Howells *et al*., 2018), and a recent study reported it in wheat (Zhang *et al.*, 2019). The A genome allele of Line 41, for example, seemed to show new, additional edits at previously intact sites, specifically three new alleles were seen in Plants 10, 38 and 82. These showed edits at positions that did not appear to be edited in the T1 Plant, suggesting that Cas9 may still have been active in these plants and that editing was still occurring. *Cas9* was shown to be still present in 8 of the 10 Line 41 T2 plants, including Plants 10, 38 and 82 (Chapter 7; section 7.2.4). Continued activity could increase the chances of effective edits being generated but adds to the difficulties of characterising the mutations that have occurred.

As mentioned in Chapter 5, this study achieved a high mutation rate, with 11 lines derived from the 14 selected T0 plants showing editing. Most of the edits characterised were deletions, with the longest deletion being 173 bp; however, there were some insertions, all of a single base pair, as well as some base pair substitutions, reflecting what was reported by Zhang *et al*. (2019) (Section 1.18). However, this chapter further highlights the problem of edit detection when using the CRISPR/Cas9 system.

This is partly because wheat is a hexaploid species, and a greater depth of NGS data is required to be able to characterise the edits effectively. The fact that wildtype alleles appeared in the T2 generation, despite not being detected in the T1 generation, and editing continued in some lines beyond the T0 generation, leads to the conclusion that several generations may be required before edited wheat lines become genetically stable.

Chapter 7

TaASN2-Knockout Phenotype and Amino Acid Analysis

7.1. Introduction

7.1.1. Asparagine in nitrogen metabolism

Asparagine, alongside glutamine, has been implicated as being critical in nitrogen metabolism in plants, with roles in transport and storage (Luo *et al*., 2017; Lea *et al*., 2007; Masclaux-Daubresse *et al.,* 2008).

Nitrogen has a critical role in plant growth and development. It is usually taken up by plant roots as either ammonium or nitrate (Xu *et al*., 2012). The nitrate that is not used directly by the roots is converted to ammonium, which is then assimilated into glutamine and either transferred directly to other parts of the plant or converted to asparagine (Xu *et al*., 2012). Asparagine has a high nitrogen to carbon ratio and is relatively inert (Lea *et al*., 2007), with high solubility and high mobility. Thus, inorganic ammonium is assimilated into asparagine, where it can be transported, recycled or stored, and this process is vital for plant development and responses to environmental stimuli (Gaufichon *et al*., 2010). Asparagine has also been implicated in the transport of nitrogen from senescing organs to growing leaves and developing seeds (Lea *et al*., 2007; Oliveira *et al*., 2001).

Some species, such as soybean (*Glycine max*), pea (*Pisum sativum*) and clover (*Trifolium* species) (Sieciechowicz *et al*., 1988) have been shown to preferentially use asparagine over glutamine, the other main nitrogen transport molecule. Environmental conditions can affect this, with asparagine levels increasing in dark-adapted pea plants (Urquhart and Joy, 1981), as a result of preferential synthesis of asparagine under carbon-limited conditions (Urquhart and Joy, 1982). Moreover, asparagine accumulates during senescence and during seed germination, as well as in response to sugar starvation and several other abiotic and biotic stimuli (Lea *et al*., 2007; Brouquisse *et al*., 1992).

7.1.2. Asparagine in the grain

In wheat grain, free asparagine accumulates in response to sulphur starvation (Curtis *et al*., 2019; Curtis *et al*., 2018b; Curtis *et al*., 2009; Granvogl *et al*., 2007; Shewry *et al*., 1983; Shewry *et al*., 2009), nitrogen fertilisation (Martinek *et al*., 2009; Claus *et al*., 2006a) and biotic stress (Curtis *et al*., 2016; Martinek *et al*., 2009). Sulphur deficiency, in particular, can lead to huge changes in free asparagine levels, with increases of up to 30-fold seen (Muttucumaru *et al*., 2006). Even under optimal conditions, there is a wide range in free asparagine concentration in the grain of different wheat varieties, with concentrations ranging from 0.708 to 11.29 mmol/kg in a UK field trial of 59 varieties, for example (Curtis *et al*., 2018b).

Free asparagine is generally one of the most abundant free amino acids in cereal grain. In an early study, for example, a maize line was shown to have a free asparagine concentration of 8 % of the total free amino acid pool (Sodek and Wilson, 1971). A later review concluded that asparagine generally accounts for up to 10 % of the free amino acids in cereal grain (Lea *et al*., 2007), and that increased to

15 % in barley supplied with excess nitrogen (Lea *et al*., 2007). A recent study of the free asparagine concentrations in a range of cereal species showed rye to have the highest concentration, 837.6 to 1535.7 mg/kg (6.33 to 11.62 mmol/kg), with oat the second highest, 510.7 to 1196.6 mg/kg (3.87 to 9.06 mmol/kg) (Žilić *et al*., 2017) and bread wheat the lowest, 430.2 to 837.4 mg/kg (3.26 to 6.34 mmol/kg). However, only a few varieties of each species were analysed in that study (Žilić *et al*., 2017), and asparagine concentration is known to be variety-dependant (Taeymans *et al*., 2004; Curtis *et al*., 2010; 2018b). In potato (*Solanum tuberosum*) tubers the concentration is even higher, with asparagine the predominant free amino acid, accounting for up to 25 % of the total free amino acids (Koch *et al*., 2003).

7.1.3. Asparagine synthetase in plant development

Despite the importance of asparagine in nitrogen metabolism, there is comparatively little research into the role of asparagine and the asparagine synthetase genes in plant development, particularly in important crop species.

Asparagine is synthesised by asparagine synthetases, encoded by asparagine synthetase genes (*ASN*s) (Chapter 3). Arabidopsis possesses three *ASN* genes, *AtASN1-3* (Lam *et al*., 1994, 1998), and these genes are expressed in response to very different stimuli (Gaufichon *et al*., 2010; Oliveira *et al*., 2002; Lam *et al*., 1998). *AtASN2* plays an important role in ammonium metabolism (Igarashi *et al*., 2009) and nitrogen assimilation and export during vegetative growth (Gaufichon *et al*., 2013, 2016). Overexpression of *AtASN1* leads to increased asparagine in the phloem, as well as higher amino acid levels in the seeds (Lam *et al*., 2003), with the *atasn1* mutation affecting seed development (Gaufichon *et al*., 2017).

There are two *ASN* genes in potato, known as *StAst1* and *StAst2* (Rommens *et al*., 2008). *StAst1* was predicted to have a role in assimilating, transporting and storing nitrogen (Chawla *et al*., 2012), similar to *AtASN1* (Lam *et al*., 2003), whilst *StAst2* was predicted to have a role in ammonium metabolism (Chawla *et al*., 2012), similar to *AtASN2* (Wong *et al*., 2004). The simultaneous silencing of both *StAst* genes through RNAi led to a 95 % reduction in free asparagine in the tubers (Rommens *et al*., 2008). Under glass, the transgenic potatoes appeared normal, with no obvious phenotype; however, the plants exhibited a deleterious phenotype when placed in the field. The *StAst1/2*-silenced plants produced small, cracked tubers, and showed a reduced yield (Chawla *et al*., 2012). When individual genes were repressed through RNAi, the *StAst1* gene was found to have an important role in asparagine synthesis in the tubers, while *StAst2* influenced more agronomic traits (Chawla *et al*., 2012). The tuber-specific silencing of *StAst1* was found to reduce the asparagine concentration of the tubers by 80 %, without negative yield phenotypes (Chawla *et al*., 2012). This shows that the free asparagine accumulating in the tubers is being synthesised *in situ* rather than imported from elsewhere in the plant (Muttucumaru *et al*., 2014a). Cracking in potato tubers has previously been shown to be

associated with stress (Jefferies and Mackerron, 1987); however, Chawla *et al*. (2012) concluded that the negative phenotype observed in their study was a direct result of the silencing of the *StAst* genes simultaneously, as no other indications of stress were noted.

The two *OsASN* genes in rice (Chapter 3) show different expression patterns, and different responses to ammonium (Nakano *et al*., 2000; Kawachi *et al*., 2002), with *OsASN1* shown to be responsible for asparagine synthesis in the roots (Ohashi *et al*., 2015). Luo *et al*. (2019) showed the importance of *OsASN1* in plant development in rice: T-DNA inserted and CRISPR/Cas9 generated *osasn1* mutants showed reduced plant height, root length, and tiller number compared with wildtype, with the relative expression of many genes involved in the asparagine metabolic pathway declining, and highlighted the importance of *OsASN1* in tiller outgrowth. Nitrogen and phosphate are known to be important in tiller outgrowth regulation (Luo *et al*., 2017); however, the concentration of nitrogen was unaffected by the loss of *OsASN1* (Luo *et al*., 2019). The targeted knock-out of *OsASN1* affected other genes in the asparagine metabolism pathway, particularly *OsASN2*, which was upregulated in the *osasn1* mutants (Luo *et al*., 2019), with the expression of genes involved in glutamine and glutamate transformation decreasing. Note that *OsASN1* is equivalent to *TaASN4* (Chapter 3).

In grain development in wheat, *TaASN3* is the most highly expressed at the early stages; however, *TaASN1* and *TaASN2* are the most highly expressed by mid-development (Curtis *et al*., 2019; Gao *et al*., 2016). *TaASN2* is expressed much more highly than *TaASN1* (Gao *et al*., 2016, Curtis *et al*., 2019), shows grain specific expression and is unique to the Triticeae (Chapter 3).

7.1.4. Chapter aims

The aim of the work described in this chapter was to investigate any phenotypic effects associated with the successful editing of *TaASN2*. In the T2 generation, the genome edited lines were compared to wildtype plants to assess the effect of mutations in the *TaASN2* gene on plant growth and development.

The T2 and T3 seeds were tested for amino acid concentration, to indicate whether genetic interventions into the *TaASN2* gene brought about a reduction in asparagine accumulation in wheat grain and, therefore, a reduction in acrylamide-forming potential.

7.2. Results

7.2.1. Phenotypic analysis of the T2 generation

7.2.1.1. Germination issues in the T2 generation

An immediate phenotype was noted in that seeds of Lines 23, 30, 41 and 59 showed poor germination rates. The seeds grew roots, but the plants failed to produce a shoot. The shoot nub that did emerge was white in colour and eventually rotted despite the excess of roots. Potentially the seed's asparagine stores are required as the initial nitrogen source, without which the seed fails to germinate, or *TaASN2* is expressed during germination and has an important role in the germination process. This needs to be researched further. The phenotype was overcome by the exogenous application of 0.1 M asparagine to the seeds during germination.

7.2.1.2. Growth phenotype

The plant growth measures (Table 7.1) were analysed statistically using a restricted maximum likelihood (REML)/linear mixed model method. The predicted means and the average standard error of differences are shown in Table 7.1. The only variable to show a significant difference ($p < 0.05$, F-test) between the edited lines and the wildtype was the 50 seed weight. Line 99 showed a predicted mean weight of 2.43 g compared with 1.68 g for the wildtype (WT), with Lines 23 and 178 showing 2.14 and 2.15 g, respectively. There was no detectable difference in any edited line to the wildtype plants for all other measures. These results, with no obvious negative phenotypic effect of the editing in *TaASN2*, although promising, will have to be confirmed in the field.

Plant height ranged from 54 cm (WT) to 93 cm (Plant 126); however, there was variation within the plant lines themselves. The biggest range in plant height was seen in the WT plants (54 – 87 cm), with the smallest range seen in Line 99 (59 – 76 cm). Line 99 produced the shortest plants, with an average tallest tiller height of 69.22 cm, and Line 126 produced the tallest, with an average value of 79.5cm. Despite being the shortest plants (59 – 76 cm), Line 99 did not produce the smallest biomass (average of 26.69 g). Line 30 had the smallest total plant biomass (23.48 g) and Line 41 had the largest (37.64 g). The average 50 seed weight of the plants also varied considerably, with the heaviest seeds seen in Line 99 at 2.196 - 3.301 g (average of 2.43 g) compared with 0.789 to 2.253 g (average of 1.68 g) for WT plants. The lightest seeds were seen in Line 126, where the highest average 50 seed weight was only 2.172 g.

Table 7.1. The predicted means of the plant lines from the restricted maximum likelihood (REML)/linear mixed model for comparing the edited lines to the wildtype. The average standard error of differences is shown for the error compared to the WT plants and for the error between the lines, and the F probability (*F pr*) is shown for each variable.

	Predicted means of plant lines								Average standard error of differences		
	126	178	23	30	41	59	99	WT	WT	Between Lines	*F pr*
Height of tallest tiller (cm)	79.81	73.58	74.43	73.37	76.78	73.20	70.00	72.60	2.89	3.30	0.16
Tiller number	10.56	7.43	9.76	7.62	11.28	11.27	10.80	8.87	1.70	1.91	0.24
Ear number	9.45	6.95	8.28	7.03	10.26	10.17	8.69	7.96	1.44	1.58	0.26
Total above ground biomass (g)	32.00	29.36	31.38	24.38	40.15	33.78	32.20	29.04	6.57	7.07	0.59
Vegetative biomass (g)	12.53	10.18	12.94	9.30	15.77	12.05	11.17	10.81	2.42	2.58	0.38
Ear weight (g)	20.00	18.87	18.30	15.46	24.48	21.62	20.49	18.26	4.27	4.64	0.70
Total seed weight (g)	14.94	14.64	14.32	12.18	18.77	16.76	17.98	13.90	3.47	3.83	0.67
Average 50 seed weight (g)	1.84	2.14	2.15	1.98	1.79	1.80	2.73	1.68	0.15	0.17	<0.001

7.2.2. T2 free amino acid analysis

7.2.2.1. Asparagine

Free asparagine concentration was measured in the seeds of the T1 plants (i.e. the T2 seeds), with five individual seeds being analysed by GC-MS for each plant. The wildtype control seeds showed a mean asparagine concentration of 1.628 mmol/kg (values ranged from 1.112 to 2.309 mmol/kg) (Figure 7.1). The wildtype-like seeds, Line 126, also showed variation in the asparagine concentration, with levels between 0.877 – 2.829 mmol/kg, and a similar mean concentration of 1.545 mmol/kg. Line 30, a triple null, did not show a reduction in asparagine concentration, with a mean concentration of 1.560 mmol/kg.

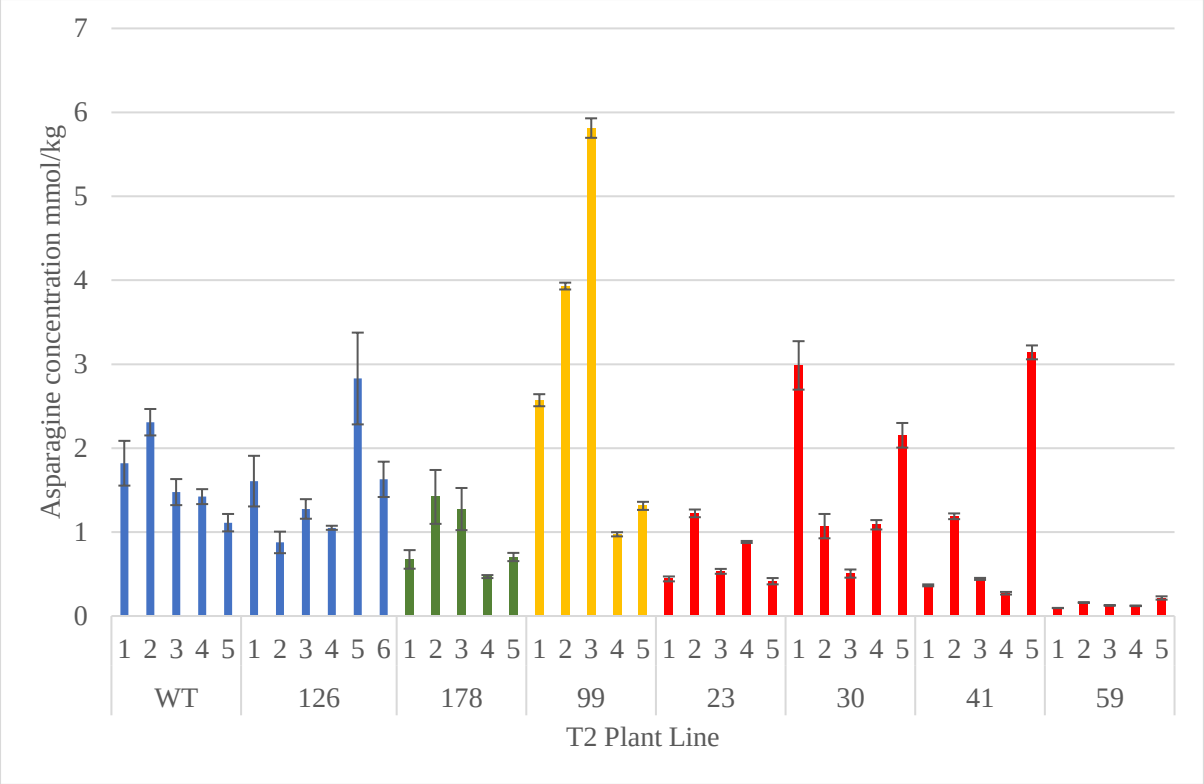


Figure 7.1. Free asparagine concentrations in the T2 grain of edited wheat (*Triticum aestivum*) cv. Cadenza seeds from: Line 178 (A genome mutant), Line 99 (AD mutant); Lines 23, 30, 41 and 59 (ABD mutants), and wildtype (WT) and Line 126 (WT-like). Individual seeds were analysed, and the graphs show the means and standard errors from three technical reps.

Line 41, a triple null line, showed a low level of asparagine in most of the seeds; however, seed 5 showed the third highest concentration seen in any plant. The overall average asparagine concentration was 1.081 mmol/kg, with seed 5 showing a concentration of 3.141 mmol/kg, nearly 12-fold higher than the lowest value seen in Line 41. Line 178, the A genome null line, showed an average asparagine concentration of only 0.908 mmol/kg, 56 % of the wildtype average. The highest concentration seen in Line 178, from seed 2, showed a concentration of 1.419 mmol/kg, just underneath wildtype average.

Line 23, a triple null line, showed a low-asparagine phenotype. The asparagine levels were variable, but even the highest concentration in Line 23, seed 2, with a concentration of 1.223 mmol/kg, was lower than the wildtype average. The average concentration of asparagine was 0.699 mmol/kg, making it the second lowest genotype for asparagine level. Line 59 exhibited a very low asparagine phenotype, with values ranging from 0.095 to 0.215 mmol/kg and an average concentration of 0.143 mmol/kg. This was less than 10% of the wildtype control line.

Line 99 showed no reduction in free asparagine concentration; instead, it showed the highest average asparagine concentration at 2.920 mmol/kg. This high average concentration was skewed by the extremely high concentration seen in seed 3 (5.813 mmol/kg) and, to a lesser extent, seed 2 (3.931 mmol/kg). The asparagine concentration in seed 3 was over 3.5-fold higher than the wildtype average, with seed 2 showing a value which was nearly 2.5 times higher. Seed 4 and 5 for Line 99 showed lower levels of asparagine, at 0.973 and 1.312 mmol/kg, respectively; values that were lower than the WT average.

7.2.2.2. Other key amino acids in asparagine metabolism

The values of glutamic acid (glutamate) also varied considerably on a seed-to-seed basis (Figure 7.2 top). The wildtype controls showed a 3-fold range in glutamate concentration and the wildtype and Line 126 seeds showed a similar level of glutamate, as expected, 0.211 - 0.737 and 0.278 - 0.643 mmol/kg, respectively, and mean values of 0.525 and 0.479 mmol/kg, respectively. Higher concentrations were seen in seeds from Line 23, showing a high of 1.721 mmol/kg with seed 2 and an average value of 1.019 mmol/kg, Line 59, which showed a high of 1.669 mmol/kg with seed 5 and an average of 1.108 mmol/kg, and Line 99, which showed the highest concentration of glutamate at 3.181 mmol/kg in seed 3 and the highest average at 1.646 mmol/kg. However, Lines 41 and 178 also showed seeds with higher glutamate concentrations, with average values of 1.057 and 0.931 mmol/kg, respectively. Only Line 30 showed an average glutamate concentration which was comparable to the wildtype/Line 126 value, with a mean value of 0.611 mmol/kg.

The aspartate concentrations also varied from seed to seed (Figure 7.2 middle), with one of the control wildtype seeds showing a very low concentration of aspartate (0.012 mmol/kg). Extremely low concentrations were seen in every plant line apart from Line 41, with seed 2 from Line 126 showing 0.073 mmol/kg, seed 1 from Line 178 showing 0.031 mmol/kg, seed 4 from Line 59 showing 0.045 mmol/kg and seed 4 from Line 99 showing 0.042 mmol/kg. No aspartate at all was detected in seeds 4 and 5 from Line 178, seeds 1 and 5 from Line 23 and seed 3 from Line 30. Line 41 showed a large variation in aspartate concentration, from 0.242 to 1.891 mmol/kg. Clearly this variation in aspartate was not genotypic in nature, as there was large variation between the Lines that could not be accounted for on the basis of the NGS data, and there was no obvious disruption in aspartate biosynthesis in the low-asparagine Line 59 seeds, with seed 5 showing the highest concentration of aspartate seen in any genotype at 2.219 mmol/kg.

The levels of glutamine in the seeds were relatively consistent (Figure 7.2 bottom) with the obvious exception of seed 3 from Line 99, which showed a glutamine concentration

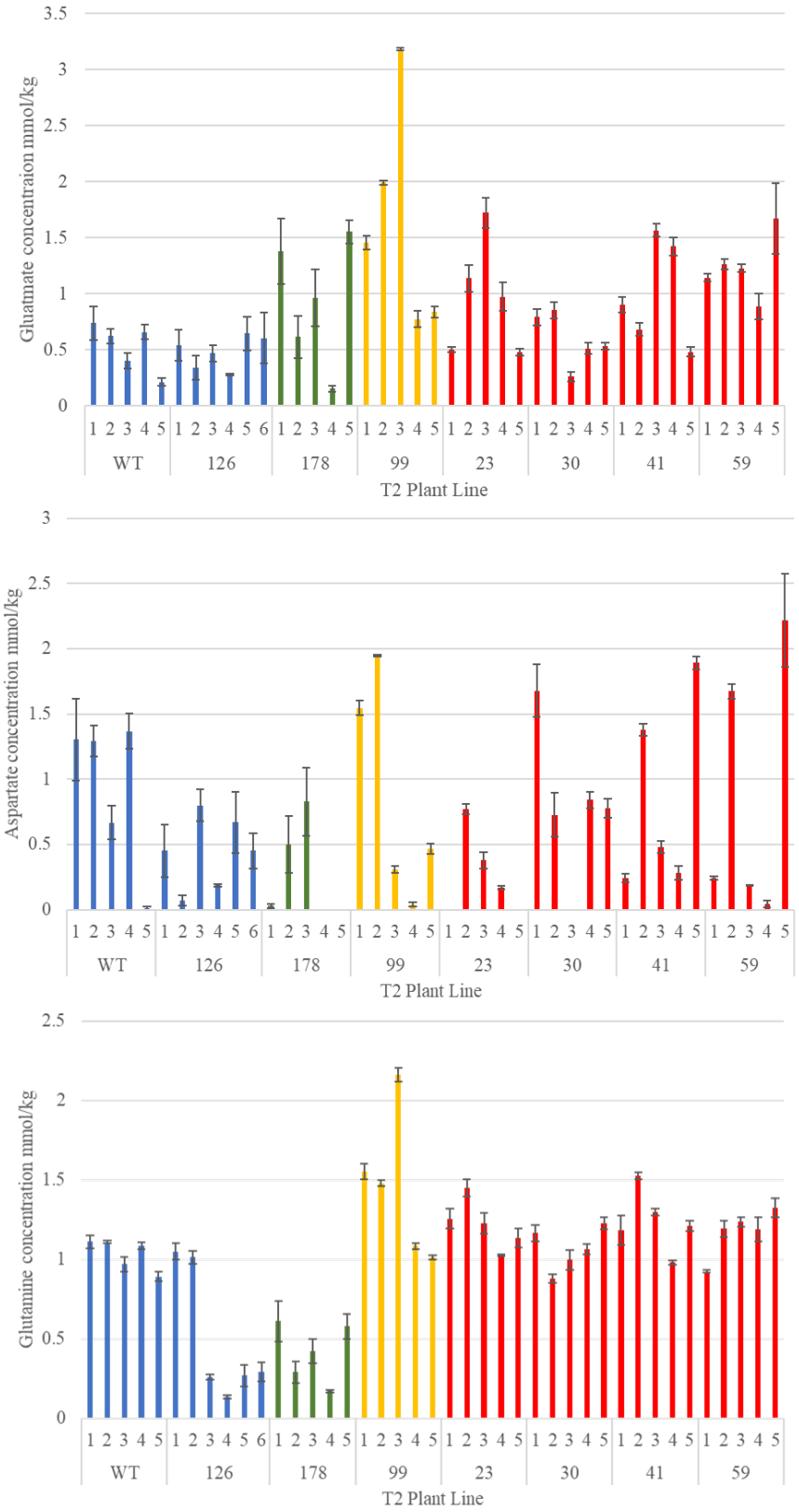


Figure 7.2. Free glutamate (top), aspartate (middle), and glutamine (bottom) concentrations in the T2 grain of edited wheat (*Triticum aestivum*) cv. Cadenza seeds from: Line 178 (A genome mutant), Line 99 (AD mutant); Lines 23, 30, 41 and 59 (ABD mutants), and wildtype (WT) and Line 126 (WT-like).

Five individual seeds were analysed per plant and the graphs show the means and standard errors from three technical reps for each seed.

10-fold higher than the wildtype controls, with 11.432 mmol/kg compared to the 0.949 - 1.175 mmol/kg seen in the wildtype plants. Two of the remaining Line 99 seeds showed a higher glutamine phenotype, with values of 1.823 and 2.339 mmol/kg; however, the other two showed levels comparable to wildtype plants. Line 23 was also consistently higher than the wildtype values, with mean glutamine values of 2.012 and 1.084 mmol/kg, respectively. Four seeds from Line 126 and two seeds from Line 178 showed low levels of glutamine.

The key amino acids in asparagine biosynthesis are synthesized in other reactions and have important biological functions which are independent of asparagine biosynthesis. Despite this, reducing asparagine synthetase gene expression has a knock-on effect on their concentrations and, therefore, the composition of the total amino acid pool.

7.2.2.3. Total free amino acid concentrations and free asparagine as a proportion of the total free amino acid pool

Some of the range in concentration of free asparagine in each line can be explained by single seeds being analysed and seed-to-seed variation, as well as genetic differences within the lines, due to segregation and, potentially, continued editing. Higher overall concentrations have previously been reported in Cadenza in a field trial (Curtis *et al*., 2018b), but free asparagine is known to accumulate in response to environmental factors.

The total amino acid concentration in the seeds was similar for most of the lines: 20.12 mmol/kg in the wildtype, 19.94 mmol/kg in Line 126, 19.16 mmol/kg in Line 178, 18.19 mmol/kg in Line 23, 20.43 mmol/kg in Line 30, 19.50 mmol/kg in Line 41 and 20.24 mmol/kg in Line 59. As a result, the lower free asparagine seen in the edited lines represented reduced asparagine specifically, rather than a reduction in total free amino acids. In the wildtype and wildtype-like seeds, asparagine represented 8.09 and 7.75 % of the total free amino acids. Line 30 showed a similar value of 7.63 % of the total amino acid pool being asparagine. Line 178 showed 4.74 %, Line 41 5.55 % and Line 23 3.84 %. In Line 59, asparagine only represented 0.71 % of the total free amino acid pool. Line 99 showed a much higher concentration of total free amino acids, 29.01 mmol/kg, and asparagine made up 10.06 % of that.

7.2.3. T3 amino acid analysis

The unedited control was so similar to wildtype plants in terms of free asparagine concentration in the T2 seed that it was not considered necessary to continue to analyse this line as a separate control. No novel editing events were detected in the T2 NGS analysis of this line either, so T3 seed were not

analysed. Line 30 was also not analysed further because it was not showing a low asparagine phenotype. Plants from the other lines were selected for T3 seed amino acid analyses based on the genetics from the T2 NGS data (Chapter 6). This reduced the number of plants that were analysed; in Line 23, for example, only the three plants that were shown to be triple nulls (plants 33, 60 and 75) were analysed.

Due to more grain being available for analysis, and the variability seen in the single seed analysis at the T2 stage, the amino acid concentrations in the T3 seeds were measured after the seeds had been ground to wholemeal flour, with the seed from a single plant being bulked together to give one batch of flour. The samples were analysed using high performance liquid chromatography instead of gas chromatography – mass spectrometry due to a change in methodology at the analytical laboratory (Chapter 2; section 2.6.2 and 2.6.3).

7.2.3.1. Asparagine

The wildtype T3 seeds (plants P1, P32 and P83) showed a much higher asparagine concentration than the T2 seeds, with the mean values ranging from 13.89 – 17.17 mmol/kg, and the average of all the wildtype plants being 15.07 mmol/kg (Figure 7.3).

Line 23 (plants P33, P60 and P75) ranged from 7.75 to 10.15 mmol/kg, with a total average for Line 23 of 8.57 mmol/kg, which is just over half the amount seen in the wildtype lines. Line 178 (P23, P35 and P51) ranged from 7.98 to 13.41 mmol/kg, with a line average of 10.24 mmol/kg, which corresponds to two thirds of the wildtype value. Only two plants were tested in Line 99, P28 and P52, which showed asparagine concentrations of 14.08 and 8.57 mmol/kg, respectively, which gives an average of 11.33 mmol/kg. This is three quarters of what is seen in the wildtype seeds. The very high levels of free asparagine seen in some of the T2 seeds from this line were not evident in the T3 generation.

There were five plants tested in Line 59, P6, P16, P26, P59 and P84, and the free asparagine concentration ranged from 3.74 to 10.43 mmol/kg, with an overall average of 7.22 mmol/kg. This is just under half the value seen in the wildtype plants, although P84 showed an asparagine concentration that was only a quarter of that seen in the wildtype plants.

Three plants were tested in Line 41 (P10, P43 and P69), and these three plants showed a huge range in asparagine concentration. Plant 69 showed a concentration of 8.57 mmol/kg, which is just over half of what is seen in the wildtype seeds and identical to the average for Line 23. However, Plants 10 and 43 showed the two highest asparagine concentrations seen in the analysis, with 19.61 and 26.12 mmol/kg respectively. This leads to a line average of 18.10 mmol/kg, which is 20 % higher than the wildtype concentration.

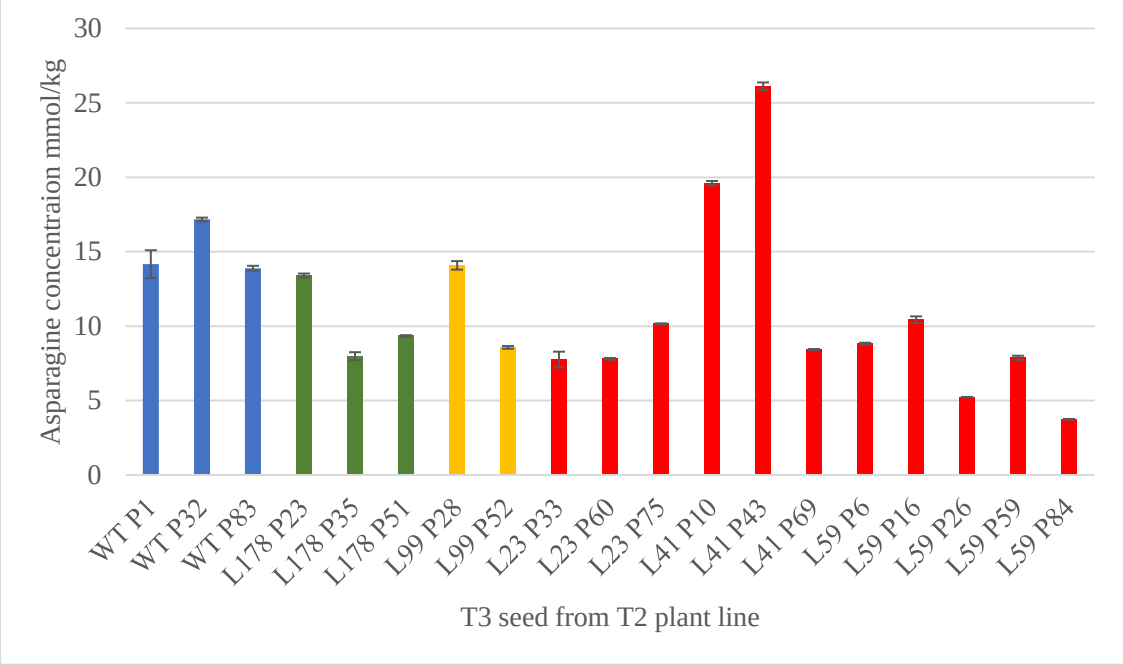


Figure 7.3. Free asparagine concentrations in wholemeal flour from the T3 grain of edited wheat (*Triticum aestivum*) cv. Cadenza plants from lines: wildtype (WT); Line 178 (A genome mutant); Line 99 (AD mutant); Lines 23, 41 and 59 (ABD mutants). The graphs show the means and standard errors from three technical reps.

An analysis of variance (ANOVA) showed the effect of line on asparagine concentration to be significant (*F pr* = 0.04). T-tests (two-sided) were then performed to compare the lines to the wildtype (WT) (Table 7.2), and the reduction in free asparagine in Line 59 was shown to be significant (*p* = 0.024, t-test); however, the T-tests were not corrected for multiple testing and are not significant when Bonferroni correction is applied. The reductions in asparagine in Lines 99, 23 and 178 compared to wildtype were not significant (Table 7.2); more data from a larger number of reps than could be generated in this experiment would be required to test the significance of the smaller reductions in free asparagine in these lines.

Table 7.2. The results of a two-way t-test to compare free asparagine concentration in wholemeal flour produced from T3 seeds of edited lines and wildtype, with the line mean, the difference to the WT mean, the standard error of the differences (SED) and the t and p values.

	Mean	WT Difference	SED	t-val	p-val
178	10.2413	-4.8316	3.442	-1.4037	0.1838
23	8.5483	-6.5246	3.442	-1.8956	0.0805
41	18.0507	2.9778	3.442	0.8651	0.4026
59	7.2164	-7.8565	3.079	-2.5516	0.0241
99	11.3246	-3.7483	3.849	-0.9738	0.3479
WT	15.0729				

7.2.3.2. Other key amino acids in asparagine metabolism

The concentration of glutamate in the flour (Figure 7.4 top) from the wildtype grain ranged from 2.98 to 3.52 mmol/kg, averaging 3.32 mmol /kg. Line 178 showed a range similar to the wildtype, from 2.68 to 3.54 mmol/kg, with an average of 3.02 mmol/kg, which is just slightly below the wildtype average. Line 23 had an average of 3.73 mmol/kg (range 2.81 – 4.52 mmol/kg). Line 99 showed glutamate concentrations of 4.48 (Plant 28) and 3.52 (Plant 52) mmol/kg, averaging 4.00 mmol/kg, about a fifth higher than the wildtype values. The glutamate concentrations in Line 41 showed a large range, from 3.64 to 6.09, averaging at 5.20 mmol/kg, over 50 % higher than the wildtype average. Line 59 also showed a high aspartate concentration, with 6.25 mmol/kg, nearly 90 % higher than the wildtype average, with the individual plant values ranging from 5.15 to 7.18 mmol/kg. The ANOVA showed the effect of line to be significant (*F pr* = >0.001) for glutamate concentration. Significant differences to the wildtype were seen for Lines 41 and 59 (Table 7.3).

Table 7.3. The results of an analysis of variance test on the effect of line on glutamate concentration (*F pr* < 0.001). The Line means and the least significant differences (LSD) at the 5 % significance level.

		Line					
		178	23	41	59	99	WT
	Line Mean	3.021	3.731	5.204	6.251	4.001	3.321
LSD	178	*	1.469	1.469	1.314	1.642	1.469
	23		*	1.469	1.314	1.642	1.469
	41			*	1.314	1.642	1.469
	59				*	1.505	1.314
	99					*	1.642
	WT						*

The wildtype concentration of aspartate in the flour of T3 seed (Figure 7.4, middle) ranged from 3.24 to 5.03 mmol/kg, with an average value of 4.11 mmol/kg. Line 178 showed an average of 4.62 mmol/kg (4.13 – 5.35 mmol/kg). The aspartate concentrations in Line 99 were 6.86 and 6.45 mmol/kg for plant 28 and 53, respectively, averaging 6.66 mmol/kg, 60 % higher than what is seen in the wildtype seeds. Concentrations in Line 23 ranged from 4.78 to 5.88 mmol/kg, with an average of 5.25 mmol/kg. Line 41 showed a higher concentration of aspartate, with values ranging from 5.77 to 10.39 mmol/kg and an average concentration of 8.12 mmol/kg. Line 59 also showed a higher aspartate concentration, with an average concentration of 8.10 mmol/kg, nearly double that of the wildtype, and values ranged from 5.86 (Plant 84) to 10.19 (Plant 59) mmol/kg. The ANOVA showed the effect of

line to be significant (*F pr* = 0.007) for aspartate concentration. Significant differences to the wildtype were seen for Lines 41 and 59 (Table 7.4).

Table 7.4. The results of an analysis of variance test of the effect of line on aspartate concentration (*F pr* = 0.007). The Line means and the least significant differences (LSD) at the 5 % significance level.

		Line					
		178	23	41	59	99	WT
	Line Mean	4.6244	5.2465	8.1208	8.1023	6.6572	4.1057
LSD	178	*	2.473	2.473	2.212	2.765	2.473
	23		*	2.473	2.212	2.765	2.473
	41			*	2.212	2.765	2.473
	59				*	2.534	2.212
	99					*	2.765
	WT						*

The concentration of glutamine in the flour (Figure 7.4, bottom) from the wildtype grain ranged from 0.47 to 0.76 mmol/kg, with an average of 0.62 mmol/kg. Line 23 showed an average glutamine concentration of 0.77 mmol/kg, comparable to the wildtype value, with the values ranging from 0.46 to 1.19 mmol/kg. Line 178 showed an average value of 0.40 mmol/kg, with concentrations ranging from 0.15 to 0.66 mmol/kg. Line 59 showed a relatively small range in glutamine concentrations, with values ranging from 1.19 to 1.67 mmol/kg. The average concentration was 1.52 mmol/kg, 2.5 times higher than the wildtype levels. Line 41 ranged from 0.53 to 2.78 mmol/kg, with an average value of 1.74 mmol/kg, nearly three-fold higher than the average wildtype concentration. In Line 99, plant 28 showed a concentration of 1.90 mmol/kg, over three times higher than the wildtype value, whereas plant 52 showed a concentration of only 0.23 mmol/kg, which was just over a third of the wildtype value, averaging 1.07 mmol/kg for the line. The ANOVA showed that there was only a marginally significant effect of line for glutamine concentration (*F pr* = 0.078).

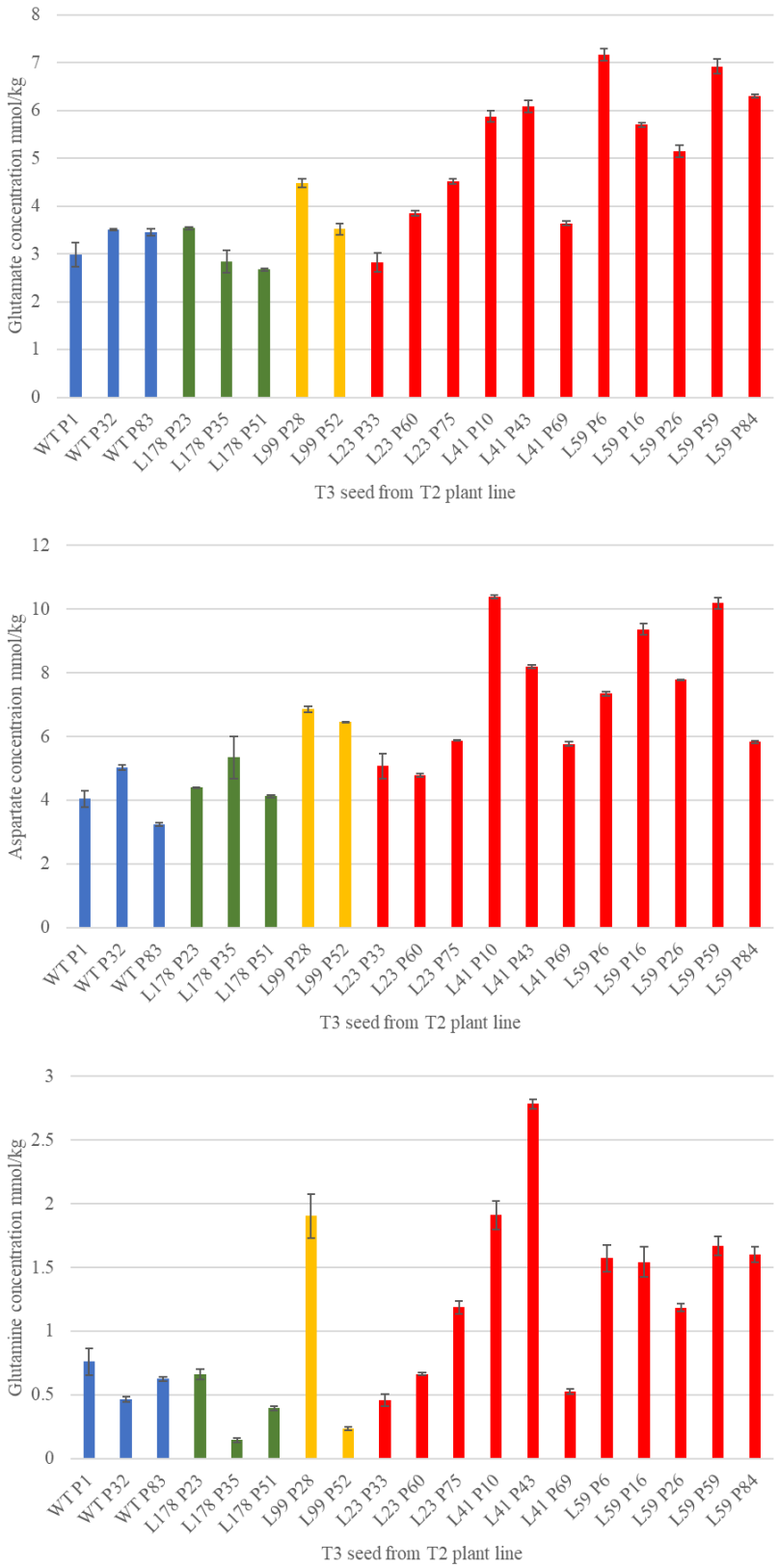


Figure 7.4. Free glutamate (top), aspartate (middle), and glutamine (bottom) concentrations in wholemeal flour from the T3 grain of edited wheat (*Triticum aestivum*) cv. Cadenza plants from lines: wildtype (WT); Line 178 (A genome mutant); Line 99 (AD mutant); Lines 23, 41 and 59 (ABD mutants). The graphs show the means and standard errors from three technical reps.

7.2.3.3. Total amino acid concentrations and free asparagine as a proportion of the free amino acid pool

For the wildtype plants, the total free amino acid concentration was 36.92 mmol/kg (34.43 to 39.73 mmol/kg). This means that asparagine made up 40.79 % of the total free amino acid pool (Figure 7.5), substantially higher than what was seen in the T2 seed.

Line 178 showed a mean total free amino acid concentration of 28.16 mmol/kg (23.93 to 35.37 mmol/kg), therefore asparagine represented 36.15 % of the total amino acids in Line 178. The mean total free amino acid concentration for Line 99 was 37.11 mmol/kg (46.67 and 27.55 mmol/kg for plants 28 and 52 respectively), with asparagine contributing 30.64 % of that. In Line 23, the average total free amino acid concentration was 29.96 mmol/kg (ranging from 25.84 to 35.88 mmol/kg), meaning that on average, asparagine represented 28.69 % of the total free amino acid pool.

Line 41 showed a higher total amino acid concentration: 51.30 mmol/kg (ranging from 30.56 to 66.32 mmol/kg). This means that the increased levels of free asparagine seen in Line 41 were in keeping with a generational increase in the concentration of free amino acids in that line and that the levels of free asparagine as a percentage of the total free amino acids (33.78 %) were not abnormally high. Line 59 averaged a total free amino acid concentration of 40.78 mmol/kg (35.57 – 48.16 mmol/kg), meaning that free asparagine only accounted for 17.42 % of the total free amino acid pool in the plants of Line 59, less than half what was seen in the wildtype seeds.

The total free amino acid pool in the T3 seed was much larger than that of the T2 seed. The average total amino acid concentration for the T2 wildtype seed was 20.12 mmol/kg, whilst the T3 seed was 36.92 mmol/kg, which is an increase of over 80 %.

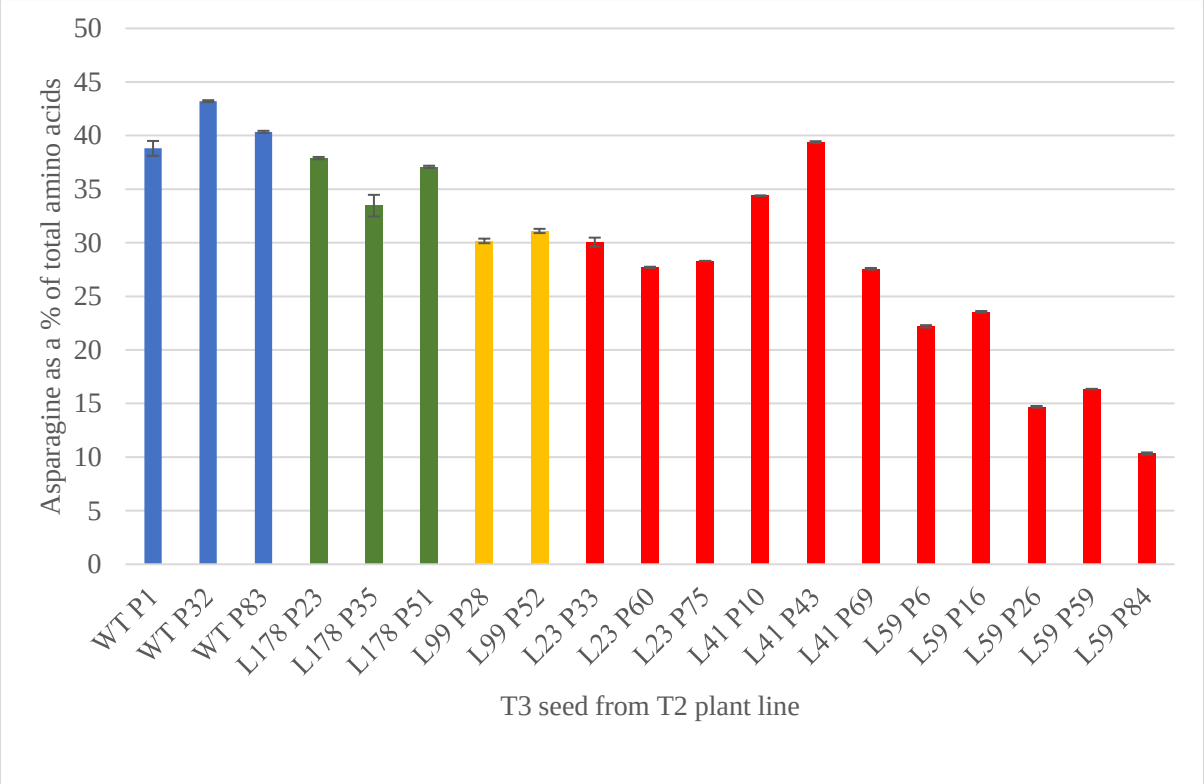


Figure 7.5. The contribution of asparagine to the total free amino acid pool in wholemeal flour produced from the T3 seed of wheat (*Triticum aestivum*) cv. Cadenza plants from lines: wildtype (WT); Line 178 (A genome mutant); Line 99 (AD mutant); Lines 23, 41 and 59 (ABD mutants). The graphs show the means and standard errors from three technical reps.

7.2.4. Analysing the T2 plants for *Cas9* integration

The presence of the *Cas9* gene in the T2 plants was assessed by PCR (Figure 7.6) to determine whether the *Cas9* gene was segregating away to produce *Cas9*-free plants. The segregating away of the *Cas9* gene is desired because it is possible that the integration of the transgene led to disruption of native genes, potentially affecting the phenotype of the plants. The presence of the *Cas9* would also raise the issue of further and novel edits, particularly as there was evidence of transgenerational editing in some of the lines (Chapter 6), meaning that genetic stability of the lines could not be achieved while the *Cas9* gene was still present.

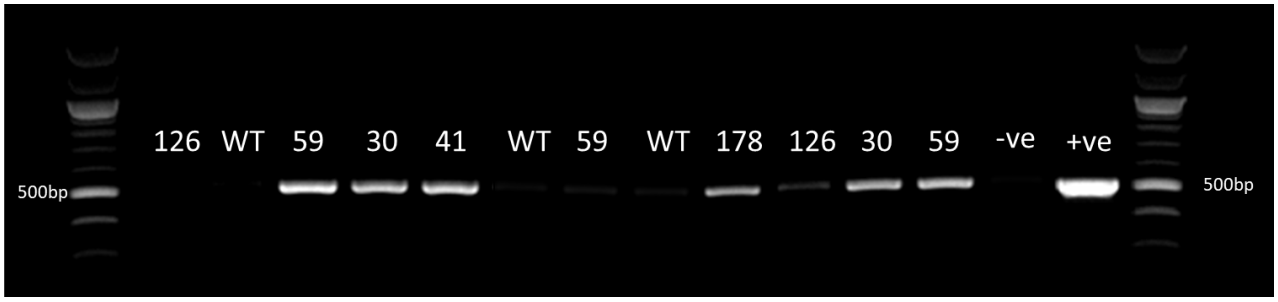


Figure 7.6. Electrophoresis gel with PCR products of part of the *Cas9* gene from DNA isolated from leaves of wheat (*Triticum aestivum*) cv. Cadenza plants. The lines the plants were derived from are indicated. Not all samples are shown. The *Cas9*-containing plasmid (pRRes217.486) was used as a control. The expected product was 557 bp and the 2-Log DNA Ladder (0.1–10.0 kbp) (NEB) was used.

The *Cas9* gene was detected in two of the Line 126 plants (Plants 3 and 51). These plants have been classified as wildtype-like due to no editing events being present in the T2 NGS data. Despite the integration of *Cas9* plants, no further editing events were detected in the T3 generation. Line 41 only showed 20 % of the plants being *Cas9*-free (Plants 15 and 29). Line 99 showed seven *Cas9*-free plants out of a total of nine plants (Plants 4, 25, 34, 47, 62, 74, 78 and 87). Line 178 showed 60 % of plants as *Cas9*-free (Plants 13, 23, 35, 42, 64 and 76). Line 23 showed integration in 80% of the plants (Plants 45 and 75 were *Cas9*-free). Line 59, the very low asparagine line, showed only one plant out of 11 to be *Cas9*-free (Plant 84), and all the ten Line 30 plants tested showed *Cas9* integration, indicating a multiple copy number of the gene.

7.3. Discussion

7.3.1. Plant growth phenotype

The genome edited *taasn2* plants showed no deleterious growth or developmental phenotype, other than a germination phenotype which was overcome by the exogenous application of asparagine. The phenotypic effects seen in rice, when *OsASN1* (equivalent to *TaASN4*, Chapter 3, section 3.2.3.11) was targeted (Luo *et al*., 2019), and in potato, when *StAst1* and *StAst2* were simultaneously targeted (Rommens *et al*., 2008), were not seen. The *osasn1* mutants showed reduced plant height, root length, and tiller number, while many other asparagine metabolism genes were affected, particularly *OsASN2*, which was upregulated in the mutant plants (Luo *et al*., 2019). It was hypothesised that this was due to the reduced free asparagine, leading to a regulatory compensation mechanism. This has not been investigated yet in these lines. Potatoes that had both the *ASN* genes silenced showed a 95 % reduction in free asparagine (Rommens *et al*., 2008). However, they produced small, cracked abnormal tubers with reduced yield in the field (Chawla *et al*., 2012).

TaASN2 is a grain-specific gene, unique to the Triticeae (Raffan and Halford, 2019; Chapter 3, section 3.3.2). As such, it is likely it can be targeted and knocked out without compromising vegetative asparagine synthetase gene expression, and without completely removing all asparagine synthetase activity from the grain (Curtis *et al*., 2019; Chapter 3, section 3.2.2). The methods employed here, targeting *TaASN2* by genome editing, may not be applicable across crop species, and the *ASN* gene family must be investigated in each individual species before species-specific intervention strategies can be devised (Chapter 3, section 3.3.5).

7.3.2. Reduced asparagine

Plants with a greatly reduced free asparagine phenotype in the grain were produced, representing the achievement of the thesis goal of producing genome-edited wheat plants with a reduced acrylamide-forming potential. The concentrations of free asparagine in the T2 seeds of Line 59 were less than 10 % those of the wildtype controls, and less than 50 % in the T3 seeds.

The proportion of the free amino acid pool represented by asparagine in the T2 seeds from Line 59 was much lower than in the corresponding wildtype seeds, with asparagine representing less than 1 % of the free amino acids. Glutamate, and even glutamine concentration, also appeared to be higher than in the wildtype seeds. Free asparagine also represented a lower proportion of the free amino acid pool in the T3 seeds in Line 59 when compared to the wildtype; however, this was greater than in the T2 seeds. The free asparagine concentrations were also reduced in Lines 23, 59 and 178 when compared to the wildtype; however, Line 59 consistently showed the lowest free asparagine concentration. This might be a product of the different edits present in the plants (Chapter 6), and the functional effect those edits will have on the TaASN2 protein encoded; however, Line 59 appeared consistently low despite the fact there are two different A genome edited alleles, resulting in the plants not being genetically identical. Line 23 was much lower than the wildtype in the T2 seeds but this was not seen in the T3 seeds; however, only three plants were looked at in the T3 seed due to segregation. No statistical significance at 5% level in the free asparagine concentration of the T3 seed but, given the marginal significance ($F\ pr = 0.078$), it is reasonable to further investigate this line. The fact that Line 41 did not show comparable reductions in free asparagine levels also raises questions. Free asparagine accumulation is heavily influenced by environmental factors and it is also possible that the other *TaASN* genes may have a role in this.

7.3.3. Comparison of the amino acid analysis of the T2 and T3 seeds

The total amino acid concentration was a lot higher in the T3 generation, with an 83.5 % increase between the T2 and T3 wildtype seeds. The free asparagine levels were also much higher in the T3 seed, both *per se* and as a percentage of the total free amino acid pool. This increase in free amino acids, particularly asparagine, has been associated with stress (Lea *et al*., 2007; Curtis *et al*., 2014b).

The T2 plants did experience stress as the seeds were developing during the heatwave in the summer of 2019, so it is likely that this increase in asparagine is a result of heat stress. As the plants are classified as GM under current EU regulations (Chapter 5, section 5.1.3.), they could not be moved from the glasshouse when daytime temperatures increased. The increase in free asparagine concentration occurred despite the lack of a functional *TaASN2* gene, of course, suggesting that one or more of the other *TaASN* family members was responsible. *TaASN1* is known to be environmentally responsive (reviewed by Raffan and Halford, 2019) and has been shown to be responsive to salt stress, osmotic stress and abscisic acid (Wang *et al*., 2005), so it could be *TaASN1* that was responsible for this stress-induced asparagine accumulation. It is an important result that despite this increase in free amino acid and free asparagine concentration, the edited lines continued to show a reduced asparagine phenotype when compared to a wildtype control.

7.3.4. Confirmation of phenotypes

No developmental phenotypes were observed in edited plant lines; however, these results need to be confirmed in the field. Stress factors are known to affect free asparagine accumulation, as is seen in the T3 seed from the T2 heat-stressed plants. Furthermore, crop management factors also affect free asparagine accumulation (reviewed in Raffan and Halford, 2019), which may overcome the genetic intervention. The *StAst1/2* RNAi potatoes showed no growth phenotype when grown under glass (Rommens *et al*., 2008); however, a deleterious growth phenotype was seen when the plants were placed in the field (Chawla *et al*., 2012). The potato tubers were small and cracked, with an overall reduced yield, Thus, it is possible that an unexpected phenotype may been seen when the *TaASN2*-knockout plants are placed in the field, re-emphasising the need for field trial testing.

Chapter 8

Conclusions and Future Work

8.1. Restatement of the project aims

The aim of this thesis was to generate wheat (*Triticum aestivum*) plants with a low acrylamide-forming potential. Acrylamide is a Group 2a human carcinogen which was discovered in cooked foodstuffs in 2002 as a food processing contaminant. The project utilised genome editing technology for the targeted knockout of asparagine synthetase 2 (*TaASN2*) in wheat, and also identified EMS-mutagenised lines with mutations in the *TaASN2* gene.

Towards this aim, the project had four objectives:

1. Analysis of the *TaASN* genes in wheat leading to the design of effective gRNAs for the targeted knock-out of *TaASN2* through CRISPR/Cas9 technology.
2. The identification of TILLING lines with mutations in *TaASN2*.
3. Generation and characterisation of genome-edited plants with mutations in *TaASN2*.
4. Analysis of the *TaASN2*-knockout phenotype for plant growth and grain composition, and for asparagine accumulation.

8.2. Achievement and discussion of specific chapter aims

8.2.1. Chapter 3

The work described in Chapter 3 aimed to characterise the cereal *TaASN* genes *in silico* by assessing the gene family structure, with a particular focus on *TaASN2* and its homologues. This would provide further understanding of how the cereal asparagine synthetase gene family evolved and the roles of specific genes within each species. A genomic analysis was performed on species from the Triticea and less closely related cereal species. Understanding the gene families may also shed light on the genetic factors that underlie the differences in free asparagine levels in the different species, as well as informing strategies for genetic interventions in the different species, including how 'transferable' the technology used in this project would be likely to be.

The analysis confirmed the four *ASN* gene groups in the Triticeae, highlighting the inconsistency in the pre-existing *ASN* gene family nomenclature, and confirmed the subdivision of *TaASN3* into *TaASN3.1* and *TaASN3.2*. The study allowed for an evolutionary pathway for the development and division of the *ASN* gene family in cereals to be proposed. The duplication of the group 3 genes (into 3.1 and 3.2) is only found in the Triticeae, although maize has a second group 4 gene, and some truncated genes. The group 2 genes, the group that *TaASN2* belongs to, was unique to the Triticeae. This shows that the genetic interventions and technology employed in wheat could not be easily transferred to cereal species such as rice, maize and sorghum, and the unique presence of the *TaASN2*

gene, with grain-specific expression, presents a singular opportunity for reducing the acrylamide-forming potential of wheat.

RNA-seq analysis was used to confirm the grain-specific expression of *TaASN2*, and whether it was the best target for genetic interventions aimed at bringing about grain-specific reduction in free asparagine concentration. It allowed for the assessment of whether additional targets would be needed for the effective reduction of asparagine accumulation. The expression analysis was also required to be able to interpret the outcomes of the genome-editing intervention, as if there had been no phenotype in the edited plants, the expression analysis would have been needed to explain why this was, in the context of the underlying molecular network.

TaASN1 and *TaASN2* were previously known to be the most highly expressed asparagine synthetase genes in wheat grain, with *TaASN2* more highly expressed than *TaASN1* (Gao *et al*., 2016). The RNA-seq analysis confirmed that *TaASN2* was the most highly expressed *TaASN* gene in the grain and showed that its expression in the embryo was likely to be the major determinant of free asparagine concentration in the grain. The study also showed the differential expression of the *TaASN2* homeologues, where the A genome homeologue was more highly expressed than the D genome homeologue, and no B genome homeologue could be identified, indicating it was either missing or was not expressed. The RNA-seq analysis confirmed that targeting *TaASN2* was logical to reduce asparagine accumulation in wheat grain. It also indicated that genetic interventions aimed at the A genome homeologue alone may lead to step changes in grain free asparagine concentration. This strategy was adopted alongside the production of total knockouts in case the loss of all three *TaASN2* homeologues led to undesirable phenotypes.

8.2.2. Chapter 4

The work described in Chapter 4 aimed to design multiple gRNAs to edit the first exon of *TaASN2*, where the glutamine binding domain is encoded, in order to impede TaASN2 protein function. Four gRNAs, gRNA1-4, were designed, each targeting all three homeologues simultaneously, and successfully placed in a single multiplexed gRNA construct, pRRes209.481.ASN2. The multiplexed gRNA gene contained the four gRNAs interspersed with tRNAs and was constructed *via* a Golden Gate assembly. The gRNAs were designed using an improved gRNA scaffold (Dang *et al*., 2015) to improve editing efficiency. The gRNAs were successfully designed to target conserved regions in the first exon of *TaASN2* in the cv. Cadenza genome (Appels *et al*., 2018). Cadenza contains three homeologues of the *TaASN2* gene, so the potential gRNA target sites were limited further than in the cv. Chinese Spring genome, which only contains A and D genome homeologues.

A protoplast trial system was proposed to test the gRNAs prior to stable transformation of wheat embryos. The aim was to test gRNA efficiency before the expensive investment in wheat embryo

transformation. Wheat cv. Cadenza leaf protoplasts were transformed with gRNA1 and gRNA2, and a *Cas9*-containing plasmid. Extracted protoplast DNA was screened for evidence of editing; however, editing could not be detected. The effectiveness of the gRNAs was difficult to judge as the protoplast test assays showed unreliable and low transformation efficiencies. Considering the later success of the designed gRNAs, it is unlikely that editing was not seen because of low editing efficiency; rather, it is likely due to the inefficient transformation and poor-quality protoplast material. A quick, efficient way to detect editing events from the background WT protoplast population would be needed for the protoplast system to be effective.

8.2.3. Chapter 5

Chapter 5 aimed to generate stably transformed genome-edited wheat plants with mutations in the *TaASN2* genes. Isolated wheat cv. Cadenza embryos were transformed *via* biolistic bombardment with the pRRes209.481.ASN2 plasmid, alongside a *Cas9*-containing construct and a construct used for selection. The T0 plants were regenerated and screened *via* PCR to identify plants carrying large-scale editing events, such as large deletions or insertions, with successful editing producing a size difference in the subsequent amplicon (Nekrasov *et al.*, 2017). However, there were a variety of smaller edits (as were seen in the Next Generation Sequencing (NGS) analysis) which would lead to loss-of-function of the glutamine-binding domain just as effectively as large deletions, for example by introducing a frameshift. Editing was not accurately assessed in the T0 generation, so plants were taken forward to the T1 generation based on survival and seed number. The presence of editing in the T0 lines was later estimated from the T1 NGS analysis, with 11 out of the 14 lines showing some form of editing.

Alongside this, a wheat TILLING population was analysed, and mutations were identified in the A, B and D genome homeologues of the *TaASN2* gene. As the TILLING lines possess stop mutations in the 2nd and 3rd exons, the glutamine-binding domains might not be compromised in the generated proteins (Xu *et al.*, 2018), but they should lack an aspartate-binding domain. The mutations are in the process of being stacked in a single line by RAGT, one of the project partners.

8.2.4. Chapter 6

Chapter 6 aimed to characterise the edits present in the T1 and T2 plant generations, using a multiplexed NGS nucleotide sequence analysis, in order to predict their functional affects and identify edited lines to progress forward. The NGS data for the two generations were very different, with the NGS of the T2 generation showing a much greater read number: the T1 analysis generated 171,510 reads in total, and the T2 analysis 5,744,714 reads. More plants were analysed in the T1 generation, leading to an average read number per plant of 969 compared with 82,067.3 for the T2 plants, which is nearly a 100-fold increase in coverage. Most of the edits characterised were deletions, as reported

by Zhang *et al*. (2019), with the longest deletion being 173 bp; however, there were some insertions as well, all of a single base pair, as well as some base pair substitutions.

Differences in the edits present were seen from the T1 to T2 plants, with only one line, Line 178, showing identical editing alleles in the T1 and T2 generations. Due to the increased number of reads per plant, the T2 analysis allowed for greater determination of the reads present, and the relatively low number of reads per plant in the T1 generation may have masked or misrepresented some of the edits present, particularly in the B genome, which was under-represented in the data. However, the identification of new genome editing events in the T2 analysis, particularly in Line 41, indicated ongoing CRISPR/Cas9 activity, a phenomenon that has been debated in wheat (Howells *et al*., 2018; Wang *et al*., 2018; Zhang *et al*., 2019). This ongoing activity could increase the chances of editing events occurring; however, it also increases the difficulty of generating stable, well-characterised lines. This chapter also highlights the difficulties of edit characterisation, particularly in a hexaploid species such as wheat, with great depth of NGS data required to be able to characterise the edits effectively.

8.2.5. Chapter 7

Chapter 7 aimed to investigate any phenotypic effects associated with the successful editing of *TaASN2*. The genome edited lines were compared to wildtype plants in the T2 generation to assess and characterise any effect of mutations in the *TaASN2* gene on plant growth and development. The genome edited *TaASN2*-knockout plants showed no obvious growth or developmental phenotype. This is likely because *TaASN2* is a grain-specific gene, unique to the Triticeae, and so it can be targeted without compromising vegetative asparagine synthetase gene expression. A germination phenotype was noted; however, this was overcome by the application of exogenous asparagine. These results are encouraging but need to be confirmed in the field.

The amino acid concentration was tested in the T2 grain (from single seeds) and T3 grain (from wholemeal flour) to assess whether the genetic interventions in the *TaASN2* gene brought about a reduction in free asparagine. Plants with greatly reduced free asparagine in the grain were produced, the concentrations of free asparagine in the T2 seeds of Line 59, in particular, being less than 10 % those of the wildtype controls, and less than 50 % in the T3 seeds. The free asparagine concentrations were also reduced in Lines 23, 59 and 178 when compared to the wildtype; however, Line 59 consistently showed the lowest free asparagine concentration. The total free amino acid concentration was a lot higher in the T3 generation than the T2 generation, which is likely associated with stress. The reduced free asparagine phenotype meant that the primary aim of the project had been achieved, but this, too, needs to be confirmed in the field.

8.3. Project conclusions

The project aimed to produce wheat material with a low acrylamide-forming potential. This aim was achieved by targeting *TaASN2* by CRISPR/Cas9, and the successful editing of the *TaASN2* gene using four gRNAs to target the first exon of the gene. The success of the editing was confirmed, and the edits were characterised using NGS nucleotide sequences analysis. Edit characterisation in the transformed plants proved difficult, with a large number of reads per plant needed for effective characterisation. Lines were identified showing reduced free asparagine accumulation in the grain in both the T2 and T3 generations. Higher free asparagine was seen in the T3 generation compared to the T2 generation, likely due to heat stress, indicating the role of stress in asparagine accumulation. No detrimental growth phenotype was associated with the mutations in *TaASN2* apart from poor germination, which could be overcome by application of asparagine solution in a spray, although the lack of other phenotypes needs to be confirmed in the field. The *TaASN2* gene was shown to be unique to the Triticeae, so the genetic interventions in this study would not be applicable to a non-Triticeae cereal species without prior investigation of the *ASN* gene family structure and expression in that species.

8.4. Future work

8.4.1. Germination experiment

A germination phenotype was noted in the T2 seed during the phenotyping experiment and this needs to be further investigated. The initial investigation will compare the very low asparagine line (line 59), using both the T2 and T3 seed, with WT plants. Line 59 will be used as it showed a low asparagine phenotype in all the T2 seeds tested, and this will serve to confirm the germination phenotype that was previously observed. The T2 and T3 seeds will both be tested as both lines lack a functional *TaASN2* gene, but the T3 seed shows a much higher concentration of free asparagine. This will allow for assessment of whether *TaASN2* expression is important in wheat grain germination or whether it is the concentration of free asparagine itself in the grain that is important. The germination phenotype might have implications for pre-harvest sprouting and malting and could lead to novel strategies for crop improvement in these areas.

Free asparagine was seen to rescue the phenotype for the T2 seeds, and this will be confirmed, although other treatments will be also investigated to determine whether asparagine is specifically needed, or whether another nitrogen source could substitute. The seeds will be germinated on filter paper in a Petri dish, which will be saturated with one of the following: asparagine, glutamine, ammonium nitrate, ammonium sulphate and the control treatment of just water. Glutamine is one of the precursors for asparagine formation and the primary organic product of nitrogen assimilation. It

will show whether an organic nitrogen source is required. Ammonium nitrate and ammonium sulphate are agricultural fertilisers and would be much more readily applicable and familiar to farmers if very low asparagine varieties were commercialised and a simple solution to the phenotype was required in the field. Ammonium nitrate is a high-nitrogen fertilizer whereas ammonium sulphate has a lower nitrogen content relative to ammonium nitrate but also contains sulphur, which may be important given the link between sulphur availability and asparagine synthetase gene expression.

Five seeds per genotype will be geminated on a single filter paper, with three dishes set up per treatment. The seeds will be allowed to germinate and grow for seven days and then analysed for shoot and root development. Should the phenotype be present, the seedlings will be frozen in liquid nitrogen and homogenised after seven days so as to be analysed for free amino acid concentration.

8.4.2. Field trial to confirm growth and low asparagine phenotypes

A field trial is planned for the growing season of 2021 to 2022. This will have to be run as a GM field trial unless regulations change in the meantime, and will be a joint project between Rothamsted Research, University of Bristol and the industrial partners who supported this studentship: AHDB, KWS, Limagrain, RAGT, Saaten Union and Syngenta. Before the field trial, the plants will be bulked up to provide enough seed, and genotyped for *Cas9* and *BAR* integration so that plants without transgenes can be taken forward.

The field trial will consist of five genotypes, comprising Line 23, Line 59, Line 178, an A-genome null TILLING line and wildtype, with five reps of each, making a total of 25 plots. The plants will be analysed for emergence, growth phenotype, grain composition and yield. Expression analysis of genes known to be involved in the asparagine biosynthesis pathway will be undertaken, including all the *TaASNs*.

Bibliography

Aach, J., Mali, P. and Church, G.M., 2014. CasFinder: Flexible algorithm for identifying specific Cas9 targets in genomes. *BioRxiv*, p.005074.

AHDB (Agriculture and Horticulture Development Board), 2014. Sulphur for Cereals and Oilseed Rape. *AHDB Information Sheet 28*, AHDB, Kenilworth, UK.

Aird, D., Ross, M.G., Chen, W.S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C. and Gnirke, A., 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology*, *12*(2), pp.1-14.

Aken, B.L., Achuthan, P., Akanni, W., Amode, M.R., Bernsdorff, F., Bhai, J., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P. and Gil, L., 2017. Ensembl 2017. *Nucleic Acids Research*, *45*(D1), pp.635-D642.

Alaux, M., Rogers, J., Letellier, T., Flores, R., Alfama, F., Pommier, C., Mohellibi, N., Durand, S., Kimmel, E., Michotey, C. and Guerche, C., 2018. Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence to wheat genetic and phenomic data. *Genome Biology*, *19*(1), p.111.

Ali, Z., Shami, A., Sedeek, K., Kamel, R., Alhabsi, A., Tehseen, M., Hassan, N., Butt, H., Kababji, A., Hamdan, S.M. and Mahfouz, M.M., 2020. Fusion of the Cas9 endonuclease and the VirD2 relaxase facilitates homology-directed repair for precise genome engineering in rice. *Communications Biology*, *3*(1), pp.1-13.

Altpeter, F., Baisakh, N., Beachy, R., Bock, R., Capell, T., Christou, P., Daniell, H., Datta, K., Datta, S., Dix, P.J. and Fauquet, C., 2005. Particle bombardment and the genetic enhancement of crops: myths and realities. *Molecular Breeding*, *15*(3), pp.305-327.

Altpeter, F., Springer, N.M., Bartley, L.E., Blechl, A.E., Brutnell, T.P., Citovsky, V., Conrad, L.J., Gelvin, S.B., Jackson, D.P., Kausch, A.P. and Lemaux, P.G., 2016. Advancing crop transformation in the era of genome editing. *The Plant Cell*, *28*(7), pp.1510-1520.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J., 1990. Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), pp.403-410.

Alvarez, M.L., Guelman, S., Halford, N.G., Lustig, S., Reggiardo, M.I., Ryabushkina, N., Schewry, P., Stein, J. and Vallejos, R.H., 2000. Silencing of HMW glutenins in transgenic wheat expressing extra HMW subunits. *Theoretical and Applied Genetics*, *100*(2), pp.319-327.

Amrein, T.M., Schönbächler, B., Rohner, F., Lukac, H., Schneider, H., Keiser, A., Escher, F. and Amadò, R., 2004. Potential for acrylamide formation in potatoes: data from the 2003 harvest. *European Food Research and Technology*, *219*(6), pp.572-578.

Anand, A., Trick, H.N., Gill, B.S. and Muthukrishnan, S., 2003. Stable transgene expression and random gene silencing in wheat. *Plant Biotechnology Journal*, *1*(4), pp.241-251.

Anders, C., Niewoehner, O., Duerst, A. and Jinek, M., 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*, *513*(7519), pp.569-573.

Appels, R., Eversole, K., Stein, N., Feuillet, C., Keller, B., Rogers, J., Pozniak, C.J., Choulet, F., Distelfeld, A., Poland, J. and Ronen, G., 2018. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, *361*(6403).

Avila-Ospina, L., Marmagne, A., Talbotec, J., Krupinska, K. and Masclaux-Daubresse, C., 2015. The identification of new cytosolic glutamine synthetase and asparagine synthetase genes in barley (Hordeum vulgare L.), and their expression during leaf senescence. *Journal of Experimental Botany*, *66*(7), pp.2013-2026.

Avni, R., Nave, M., Barad, O., Baruch, K., Twardziok, S.O., Gundlach, H., Hale, I., Mascher, M., Spannagl, M., Wiebe, K. and Jordan, K.W., 2017. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science*, *357*(6346), pp.93-97.

Bae, S., Park, J. and Kim, J.-S., 2014. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics*, *30*, pp.1473–1475.

Baena-González, E. and Sheen, J., 2008. Convergent energy and stress signaling. *Trends in Plant Science*, *13*(9), pp.474-482.

Baena-González, E., Rolland, F., Thevelein, J.M. and Sheen, J., 2007. A central integrator of transcription networks in plant stress and energy signalling. *Nature*, *448*(7156), pp.938-942.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. and Horvath, P., 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, *315*(5819), pp.1709-1712.

Bauer, E., Schmutzer, T., Barilar, I., Mascher, M., Gundlach, H., Martis, M.M., Twardziok, S.O., Hackauf, B., Gordillo, A., Wilde, P. and Schmidt, M., 2017. Towards a whole-genome sequence for rye (*Secale cereale* L.). *The Plant Journal*, *89*(5), pp.853-869.

Baute, J., Herman, D., Coppens, F., De Block, J., Slabbinck, B., Dell'Acqua, M., Pè, M.E., Maere, S., Nelissen, H. and Inzé, D., 2015. Correlation analysis of the transcriptome of growing leaves with mature leaf parameters in a maize RIL population. *Genome Biology*, *16*(1), pp.168.

Beland, F.A., Mellick, P.W., Olson, G.R., Mendoza, M.C., Marques, M.M. and Doerge, D.R., 2013. Carcinogenicity of acrylamide in B6C3F1 mice and F344/N rats from a 2-year drinking water exposure. *Food and Chemical Toxicology*, *51*, pp.149-159.

Beland, F.A., Olson, G.R., Mendoza, M.C., Marques, M.M. and Doerge, D.R., 2015. Carcinogenicity of glycidamide in B6C3F1 mice and F344/N rats from a two-year drinking water exposure. *Food and Chemical Toxicology*, *86*, pp.104-115.

Belhaj, K., Chaparro-Garcia, A., Kamoun, S. and Nekrasov, V., 2013. Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. *Plant Methods*, *9*(1), pp.39.

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Patron, N.J. and Nekrasov, V., 2015. Editing plant genomes with CRISPR/Cas9. *Current Opinion in Biotechnology*, *32*, pp.76-84.

Bell, C.C., Magor, G.W., Gillinder, K.R. and Perkins, A.C., 2014. A high-throughput screening strategy for detecting CRISPR-Cas9 induced mutations using next-generation sequencing. *BMC Genomics*, *15*(1), pp.1-7.

Bennett, M.D. and Leitch, I.J., 1995. Nuclear DNA amounts in angiosperms. *Annals of Botany*, pp.113-176.

Bergmark E., 1997. Hemoglobin adducts of acrylamide and acrylonitrile in laboratory workers, smokers and nonsmokers. *Chemical Research in Toxicology*, *10*, pp.78-84.

Bergmark E., Calleman C.J., He F., Costa L.G., 1993. Hemoglobin adducts in humans occupationally exposed to acrylamide. *Toxicology and Applied Pharmacology*, *120*, pp.45-54.

Bermudo, E., Moyano, E., Puignou, L. and Galceran, M.T., 2006. Determination of acrylamide in foodstuffs by liquid chromatography ion-trap tandem mass-spectrometry using an improved clean-up procedure. *Analytica Chimica Acta*, *559*(2), pp.207-214.

Besaratinia, A. and Pfeifer, G.P., 2003. Weak yet distinct mutagenicity of acrylamide in mammalian cells. *Journal of the National Cancer Institute*, *95*(12), pp.889-896.

Besaratinia, A. and Pfeifer, G.P., 2004. Genotoxicity of acrylamide and glycidamide. *Journal of the National Cancer Institute*, *96*(13), pp.1023-1029.

Betts, M.J. and Russell, R.B., 2003. Amino acid properties and consequences of substitutions. *Bioinformatics for Geneticists*, *317*, p.289.

Bevan, M.W., Flavell, R.B. and Chilton, M.D., 1983. A chimaeric antibiotic resistance gene as a selectable marker for plant cell transformation. *Nature*, *304*(5922), pp.184-187.

Bewg, W.P., Ci, D. and Tsai, C.J., 2018. Genome editing in trees: From multiple repair pathways to long-term stability. *Frontiers in Plant Science*, *9*, pp.1732.

Bian, X., Tyrrell, S. and Davey, R.P., 2017. The Grassroots life science data infrastructure.

Bibikova, M., Golic, M., Golic, K.G. and Carroll, D., 2002. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics*, *161*(3), pp.1169-1175.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. and Bonas, U., 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, *326*(5959), pp.1509-1512.

Bonnefoy, N. and Fox, T.D., 2007. Directed alteration of Saccharomyces cerevisiae mitochondrial DNA by biolistic transformation and homologous recombination. *In Mitochondria*. pp.153-166. Humana Press.

Borrill, P., Harrington, S.A. and Uauy, C., 2019. Applying the latest advances in genomics and phenomics for trait discovery in polyploid wheat. *The Plant Journal*, *97*(1), pp.56-72.

Borrill, P., Ramirez-Gonzalez, R. and Uauy, C., 2016. expVIP: a customizable RNA-seq data analysis and visualization platform. *Plant Physiology*, *170*(4), pp.2172-2186.

Bortesi, L. and Fischer, R., 2015. The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnology Advances*, *33*(1), pp.41-52.

Bradshaw, J.E., Hackett C.A., Pande B., Waugh R. and Bryan G.J., 2008. QTL mapping of yield, agronomic and quality traits in tetraploid potato (*Solanum tuberosum* subsp. *tuberosum*). *Theoretical and Applied Genetics*, *116*, pp.193-211.

Breitler, J.C., Dechamp, E., Campa, C., Rodrigues, L.A.Z., Guyot, R., Marraccini, P. and Etienne, H., 2018. CRISPR/Cas9-mediated efficient targeted mutagenesis has the potential to accelerate the domestication of *Coffea canephora*. *Plant Cell, Tissue and Organ Culture (PCTOC)*, *134*(3), pp.383-394.

Brooks, C., Nekrasov, V., Lippman, Z.B. and Van Eck, J., 2014. Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. *Plant Physiology*, *166*(3), pp.1292-1297.

Brouquisse, R., James, F., Pradet, A. and Raymond, P., 1992. Asparagine metabolism and nitrogen distribution during protein degradation in sugar-starved maize root tips. *Planta*, *188*(3), pp.384-395.

Byrne, E.H., Prosser, I., Muttucumaru, N., Curtis, T.Y., Wingler, A., Powers, S. and Halford, N.G., 2012. Overexpression of GCN2-type protein kinase in wheat has profound effects on free

amino acid concentration and gene expression. *Plant Biotechnology Journal*, *10*(3), pp.328-340.

Cai, C.Q., Doyon, Y., Ainley, W.M., Miller, J.C., DeKelver, R.C., Moehle, E.A., Rock, J.M., Lee, Y.L., Garrison, R., Schulenberg, L. and Blue, R., 2009. Targeted transgene integration in plant cells using designed zinc finger nucleases. *Plant Molecular Biology*, *69*(6), pp.699-709.

Camerlengo, F., Frittelli, A., Sparks, C., Doherty, A., Martignago, D., Larré, C., Lupi, R., Sestili, F. and Masci, S., 2020. CRISPR-Cas9 multiplex editing of the α-amylase/trypsin inhibitor genes to reduce allergen proteins in durum wheat. *Frontiers in Sustainable Food Systems*, *4*, p.104.

Chang, Y.M., Liu, W.Y., Shih, A.C.C., Shen, M.N., Lu, C.H., Lu, M.Y.J., Yang, H.W., Wang, T.Y., Chen, S.C.C., Chen, S.M. and Li, W.H., 2012. Characterizing regulatory and functional differentiation between maize mesophyll and bundle sheath cells by transcriptomic analysis. *Plant Physiology*, *160*(1), pp.165-177.

Chapman, J.A., Mascher, M., Buluç, A., Barry, K., Georganas, E., Session, A., Strnadova, V., Jenkins, J., Sehgal, S., Oliker, L. and Schmutz, J., 2015. A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biology*, *16*(1), p.26.

Char, S.N., Neelakandan, A.K., Nahampun, H., Frame, B., Main, M., Spalding, M.H., Becraft, P.W., Meyers, B.C., Walbot, V., Wang, K. and Yang, B., 2017. An Agrobacterium-delivered CRISPR/Cas9 system for high-frequency targeted mutagenesis in maize. *Plant Biotechnology Journal*, *15*(2), pp.257-268.

Chawla, R., Shakya, R. and Rommens, C.M., 2012. Tuber-specific silencing of asparagine synthetase-1 reduces the acrylamide-forming potential of potatoes grown in the field without affecting tuber shape and yield. *Plant Biotechnology Journal*, *10*(8), pp.913-924.

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S. and Huang, B., 2013. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, *155*(7), pp.1479-1491.

Chen, J., Strieder, N., Krohn, N.G., Cyprys, P., Sprunck, S., Engelmann, J.C. and Dresselhaus, T., 2017. Zygotic genome activation occurs shortly after fertilization in maize. *The Plant Cell*, *29*(9), pp.2106-2125.

Chen, W.P., Gu, X., Liang, G.H., Muthukrishnan, S., Chen, P.D., Liu, D.J. and Gill, B.S., 1998. Introduction and constitutive expression of a rice chitinase gene in bread wheat using biolistic bombardment and the bar gene as a selectable marker. *Theoretical and Applied Genetics*, *97*(8), pp.1296-1306.

Cheng, M., Fry, J.E., Pang, S., Zhou, H., Hironaka, C.M., Duncan, D.R., Conner, T.W. and Wan, Y., 1997. Genetic transformation of wheat mediated by *Agrobacterium tumefaciens*. *Plant Physiology*, *115*(3), pp.971-980.

Cheng, X., Li, G., Manzoor, M.A., Wang, H., Abdullah, M., Su, X., Zhang, J., Jiang, T., Jin, Q., Cai, Y. and Lin, Y., 2019. In silico genome-wide analysis of Respiratory Burst Oxidase Homolog (*RBOH*) family genes in five fruit-producing trees, and potential functional analysis on lignification of stone cells in Chinese white pear. *Cells*, *8*(6), p.520.

Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S. and Kim, J.S., 2014. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Research*, *24*(1), pp.132-141.

Choulet, F., Alberti, A., Theil, S., Glover, N., Barbe, V., Daron, J., Pingault, L., Sourdille, P., Couloux, A., Paux, E. and Leroy, P., 2014. Structural and functional partitioning of bread wheat chromosome 3B. *Science*, *345*(6194).

Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J. and Voytas, D.F., 2010. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, *186*(2), pp.757-761.

Christou, P., 1996. Transformation technology. *Trends in Plant Science*, *1*(12), pp.423-431.

Christou, P., McCabe, D.E. and Swain, W.F., 1988. Stable transformation of soybean callus by DNA-coated gold particles. *Plant Physiology*, *87*(3), pp.671-674.

Clasen, B.M., Stoddard, T.J., Luo, S., Demorest, Z.L., Li, J., Cedrone, F., Tibebu, R., Davison, S., Ray, E.E., Daulhac, A. and Coffman, A., 2016. Improving cold storage and processing traits in potato through targeted gene knockout. *Plant Biotechnology Journal*, *14*(1), pp.169-176.

Claus, A., Schreiter, P., Weber, A., Graeff, S., Herrmann, W., Claupein, W., Schieber, A. and Carle, R., 2006a. Influence of agronomic factors and extraction rate on the acrylamide contents in yeast-leavened breads. *Journal of Agricultural and Food Chemistry*, *54*(23), pp.8968-8976.

Claus A., Weisz G.M., Schieber A. and Carle R., 2006b. Pyrolytic acrylamide formation from purified wheat gluten and gluten-supplemented wheat bread rolls. *Molecular Nutrition and Food Research, 50*, pp.87-93.

Clouse, J.W., Adhikary, D., Page, J.T., Ramaraj, T., Deyholos, M.K., Udall, J.A., Fairbanks, D.J., Jellen, E.N. and Maughan, P.J., 2016. The amaranth genome: genome, transcriptome, and physical map assembly. *The Plant Genome*, *9*(1), pp.1-14.

Confraria, A., Martinho, C.S.D.S., Elias, A., Rubio-Somoza, I. and Baena-González, E., 2013. miRNAs mediate SnRK1-dependent energy signaling in Arabidopsis. *Frontiers in Plant Science*, *4*, pp.197.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. and Zhang, F., 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science*, *339*(6121), pp.819-823.

CONTAM Panel (European Food Safety Authority Panel on Contaminants in the Food Chain), 2015. Scientific opinion on acrylamide in food. *EFSA Journal*, *13*, 4104.

Curci, P.L., Bergès, H., Marande, W., Maccaferri, M., Tuberosa, R. and Sonnante, G., 2018. Asparagine synthetase genes (*AsnS1* and *AsnS2*) in durum wheat: structural analysis and expression under nitrogen stress. *Euphytica*, *214*(2), p.36.

Curtis, T., Halford, N.G., Powers, S.J., McGrath, S.P. and Zazzeroni, R., 2014b. Home Grown Cereals Authority Project Report No. 525. Agriculture and Horticulture Development Board, Stoneleigh Park, Kenilworth, UK, 2014a.

Curtis, T.Y., Bo, V., Tucker, A. and Halford, N.G., 2018a. Construction of a network describing asparagine metabolism in plants and its application to the identification of genes affecting asparagine metabolism in wheat under drought and nutritional stress. *Food and Energy Security*, *7*, pp.e00126.

Curtis, T.Y., Muttucumaru, N., Shewry, P.R., Parry, M.A., Powers, S.J., Elmore, J.S., Mottram, D.S., Hook, S. and Halford, N.G., 2009. Effects of genotype and environment on free amino acid levels in wheat grain: implications for acrylamide formation during processing. *Journal of Agricultural and Food Chemistry*, *57*(3), pp.1013-1021.

Curtis, T.Y., Postles, J. and Halford, N.G., 2014a. Reducing the potential for processing contaminant formation in cereal products. *Journal of Cereal Science*, *59*, pp.382-392.

Curtis, T.Y., Powers, S.J., Balagiannis, D., Elmore, J.S., Mottram, D.S., Parry, M.A., Rakszegi, M., Bedo, Z., Shewry, P.R. and Halford, N.G., 2010. Free amino acids and sugars in rye grain:

implications for acrylamide formation. *Journal of Agricultural and Food Chemistry*, *58*(3), pp.1959-1969.

Curtis, T.Y., Powers, S.J. and Halford, N.G., 2016. Effects of fungicide treatment on free amino acid concentration and acrylamide-forming potential in wheat. *Journal of Agricultural and Food Chemistry*, *64*(51), pp.9689-9696.

Curtis, T.Y., Powers, S.J., Wang, R. and Halford, N.G., 2018b. Effects of variety, year of cultivation and sulphur supply on the accumulation of free asparagine in the grain of commercial wheat varieties. *Food Chemistry*, *239*, pp.304-313.

Curtis, T.Y., Raffan, S., Wan, Y., King, R., Gonzalez-Uriarte, A. and Halford, N.G., 2019. Contrasting gene expression patterns in grain of high and low asparagine wheat genotypes in response to sulphur supply. *BMC genomics*, *20*(1), p.628.

Dahlem, T.J., Hoshijima, K., Jurynec, M.J., Gunther, D., Starker, C.G., Locke, A.S., Weis, A.M., Voytas, D.F. and Grunwald, D.J., 2012. Simple methods for generating and detecting locus-specific mutations induced with TALENs in the zebrafish genome. *PLoS Genet*, *8*(8), p.e1002861.

Dai, S., Zheng, P., Marmey, P., Zhang, S., Tian, W., Chen, S., Beachy, R.N. and Fauquet, C., 2001. Comparative analysis of transgenic rice plants obtained by Agrobacterium-mediated transformation and particle bombardment. *Molecular Breeding*, *7*(1), pp.25-33.

Dang, Y., Jia, G., Choi, J., Ma, H., Anaya, E., Ye, C., Shankar, P. and Wu, H., 2015. Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. *Genome Biology*, *16*(1), pp.280.

Daniell, H., Vivekananda, J., Nielsen, B.L., Ye, G.N., Tewari, K.K. and Sanford, J.C., 1990. Transient foreign gene expression in chloroplasts of cultured tobacco cells after biolistic delivery of chloroplast vectors. *Proceedings of the National Academy of Sciences*, *87*(1), pp.88-92.

Davidson, R.M., Gowda, M., Moghe, G., Lin, H., Vaillancourt, B., Shiu, S.H., Jiang, N. and Robin Buell, C., 2012. Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution. *The Plant Journal*, *71*(3), pp.492-502.

De Paola, E.L., Montevecchi, G., Masino, F., Garbini, D., Barbanera, M. and Antonelli, A., 2017. Determination of acrylamide in dried fruits and edible seeds using QuEChERS extraction and LC separation with MS detection. *Food Chemistry*, *217*, pp.191-195.

Decypher, Timelogic, n.d. Decypher: Algorithm and Feature Index. [Online] Available at: http://decypher1.rothamsted.ac.uk/decypher/cgi-bin/docfilter?file=/decypher/userindex.html [Accessed January 2017].

Demeke, T., Hucl, P., Båga, M., Caswell, K., Leung, N. and Chibbar, R.N., 1999. Transgene inheritance and silencing in hexaploid spring wheat. *Theoretical and Applied Genetics*, 99(6), pp.947-953.

D'Mello, J.F. ed., 2015. *Amino acids in higher plants*. CABI.

Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J. and Root, D.E., 2014. Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. *Nature Biotechnology*, *32*(12), pp.1262-1267.

Doyle, E.L., Booher, N.J., Standage, D.S., Voytas, D.F., Brendel, V.P., VanDyk, J.K. and Bogdanove, A.J., 2012. TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Research*, *40*(W1), pp.117-122.

Draper, J., Mur, L.A., Jenkins, G., Ghosh-Biswas, G.C., Bablak, P., Hasterok, R. and Routledge, A.P., 2001. Brachypodium distachyon. A new model system for functional genomics in grasses. *Plant Physiology*, *127*(4), pp.1539-1555.

Dubcovsky, J. and Dvorak, J., 2007. Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science*, *316*(5833), pp.1862-1866.

Duff, S.M., Qi, Q., Reich, T., Wu, X., Brown, T., Crowley, J.H. and Fabbri, B., 2011. A kinetic comparison of asparagine synthetase isozymes from higher plants. *Plant Physiology and Biochemistry*, *49*(3), pp.251-256.

Dvořák, J., Terlizzi, P.D., Zhang, H.B. and Resta, P., 1993. The evolution of polyploid wheats: identification of the A genome donor species. *Genome*, *36*(1), pp.21-31.

EFSA (European Food Safety Authority), 2008. EFSA's 11th Scientific Colloquium – Acrylamide carcinogenicity – New evidence in relation to dietary exposure.

EFSA (European Food Safety Authority), 2009. Results on the monitoring of acrylamide levels in food. *EFSA Scientific Report*, *285*, 1-26.

EFSA (European Food Safety Authority), 2010. Results on acrylamide levels in food from monitoring year 2008. Scientific report of EFSA. *EFSA Journal*, *8*, 1599.

EFSA (European Food Safety Authority), 2011. Results on acrylamide levels in food from monitoring years 2007-2009 and exposure assessment. *EFSA Journal*, *9*, 2133.

EFSA (European Food Safety Authority), 2012. Update on acrylamide levels in food from monitoring years 2007 to 2010. *EFSA Journal*, *10*, 2938.

Eisenbrand, G., 2020. Revisiting the evidence for genotoxicity of acrylamide (AA), key to risk assessment of dietary AA exposure. *Archives of Toxicology*.

Engler, C., Kandzia, R. and Marillonnet, S., 2008. A one pot, one step, precision cloning method with high throughput capability. *PloS One*, *3*(11).

European Commission, 1993. Council Regulation (EEC) No 315/93 of 8 February 1993 laying down Community procedures for contaminants in food. European Commission, Brussels.

European Commission, 2010. Commission recommendation of 2 June 2010 on the monitoring of acrylamide levels in food. European Commission, Brussels.

European Commission, 2011. Commission recommendation of 10 January 2011 on investigations into the levels of acrylamide in food. European Commission, Brussels.

European Commission, 2013. Commission recommendation of 8 November 2013 on investigations into the levels of acrylamide in food. European Commission, Brussels.

European Commission, 2017. Commission Regulation (EU) 2017/2158 of 20 November 2017 establishing mitigation measures and Benchmark Levels for the reduction of the presence of acrylamide in food. European Commission, Brussels.

European Parliament (2017)

(http://www.emeeting.europarl.europa.eu/emeeting/committee/agenda/201701/ENVI?meeting=ENVI-2017-0130_1&session=01-31-14-00).

Exon, J.H., 2006. A review of the toxicology of acrylamide. *Journal of Toxicology and Environmental Health, Part B*, *9*(5), pp.397-412.

Fan, D., Liu, T., Li, C., Jiao, B., Li, S., Hou, Y. and Luo, K., 2015. Efficient CRISPR/Cas9-mediated targeted mutagenesis in Populus in the first generation. *Scientific Reports*, *5*, p.12217.

Fauser, F., Schiml, S. and Puchta, H., 2014. Both CRISPR/C as-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. *The Plant Journal*, *79*(2), pp.348-359.

Felsenstein, J., 1985. Phylogenies and the comparative method. *The American Naturalist*, *125*(1), pp.1-15.

Feng, C., Yuan, J., Wang, R., Liu, Y., Birchler, J.A. and Han, F., 2016. Efficient targeted genome modification in maize using CRISPR/Cas9 system. *Journal of Genetics and Genomics*, *43*(1), pp.37-43.

Feng, Z., Mao, Y., Xu, N., Zhang, B., Wei, P., Yang, D.L., Wang, Z., Zhang, Z., Zheng, R., Yang, L. and Zeng, L., 2014. Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proceedings of the National Academy of Sciences*, *111*(12), pp.4632-4637.

Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D.L., Wei, P., Cao, F., Zhu, S., Zhang, F., Mao, Y. and Zhu, J.K., 2013. Efficient genome editing in plants using a CRISPR/Cas system. *Cell Research*, *23*(10), pp.1229-1232.

Food and Agriculture Organization of the United Nations, 2001. Chlorpropham: Toxicological evaluation. In: Pesticide residues in food - 2000. Report of the Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and the WHO Core Assessment Group. *FAO Plant Production and Protection Paper*, *163*.

Food and Agriculture Organization of the United Nations, 2003. Trade Reforms and Food Security: Conceptualizing the Linkages. Food and Agriculture Organization of the United Nations, Rome.

Food and Drug Administration, 2016. Guidance for Industry, Acrylamide in Foods. Food and Drug Administration, College Park.

Food Standards Agency, 2017. https://www.food.gov.uk/news-updates/news/2017/15890/reduce-acrylamide-consumption.

FoodDrinkEurope, 2019. Acrylamide Toolbox 2019. FoodDrinkEurope, Brussels.

Fraley, R.T., Rogers, S.G., Horsch, R.B., Sanders, P.R., Flick, J.S., Adams, S.P., Bittner, M.L., Brand, L.A., Fink, C.L., Fry, J.S. and Galluppi, G.R., 1983. Expression of bacterial genes in plant cells. *Proceedings of the National Academy of Sciences*, *80*(15), pp.4803-4807.

Frame, B.R., Zhang, H., Cocciolone, S.M., Sidorenko, L.V., Dietrich, C.R., Pegg, S.E., Zhen, S., Schnable, P.S. and Wang, K., 2000. Production of transgenic maize from bombarded type II callus: effect of gold particle size and callus morphology on transformation efficiency. *In Vitro Cellular & Developmental Biology-Plant*, 36(1), pp.21-29.

Friedman, M., 2003. Chemistry, biochemistry, and safety of acrylamide. A review. *Journal of Agricultural and Food Chemistry*, *51*(16), pp.4504-4526.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K. and Sander, J.D., 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology*, *31*(9), pp.822-826.

Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M. and Joung, J.K., 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature Biotechnology*, *32*(3), pp.279-284.

Gao, J., Wang, G., Ma, S., Xie, X., Wu, X., Zhang, X., Wu, Y., Zhao, P. and Xia, Q., 2015. CRISPR/Cas9-mediated targeted mutagenesis in Nicotiana tabacum. *Plant Molecular Biology*, *87*(1-2), pp.99-110.

Gao, R., Curtis, T.Y., Powers, S.J., Xu, H., Huang, J. and Halford, N.G., 2016. Food safety: structure and expression of the asparagine synthetase gene family of wheat. *Journal of Cereal Science*, *68*, pp.122-131.

Gardiner, L.J., Wingen, L.U., Bailey, P., Joynson, R., Brabbs, T., Wright, J., Higgins, J.D., Hall, N., Griffiths, S., Clavijo, B.J. and Hall, A., 2019. Analysis of the recombination landscape of hexaploid bread wheat reveals genes controlling recombination and gene conversion frequency. *Genome Biology*, *20*(1), p.69.

Garneau, J.E., Dupuis, M.È., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H. and Moineau, S., 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, *468*(7320), pp.67-71.

Gaufichon, L., Marmagne, A., Belcram, K., Yoneyama, T., Sakakibara, Y., Hase, T., Grandjean, O., Clément, G., Citerne, S., Boutet-Mercey, S. and Masclaux-Daubresse, C., 2017. ASN 1-encoded asparagine synthetase in floral organs contributes to nitrogen filling in Arabidopsis seeds. *The Plant Journal*, *91*(3), pp.371-393.

Gaufichon, L., Masclaux-Daubresse, C., Tcherkez, G., Reisdorf-Cren, M., Sakakibara, Y., Hase, T., Clément, G., Avice, J.C., Grandjean, O., Marmagne, A. and Boutet-Mercey, S., 2013. Arabidopsis thaliana ASN2 encoding asparagine synthetase is involved in the control of nitrogen assimilation and export during vegetative growth. *Plant, Cell & Environment*, *36*(2), p.328.

Gaufichon, L., Reisdorf-Cren, M., Rothstein, S.J., Chardon, F. and Suzuki, A., 2010. Biological functions of asparagine synthetase in plants. *Plant Science*, *179*(3), pp.141-153.

Gaufichon, L., Rothstein, S.J. and Suzuki, A., 2016. Asparagine metabolic pathways in Arabidopsis. *Plant and Cell Physiology*, *57*(4), pp.675-689.

Gelvin, S.B., 2003. Agrobacterium-mediated plant transformation: the biology behind the "gene-jockeying" tool. *Microbiology and Molecular Biology Reviews*, *67*(1), pp.16-37.

Gene Codes Corporation, Ann Arbor, MI USA, n.d. Sequencher® version 5.4.6 DNA sequence analysis software. [Online] Available at: http://www.genecodes.com.

Ghanayem, B.I., McDaniel, L.P., Churchwell, M.I., Twaddle, N.C., Snyder, R., Fennell, T.R. and Doerge, D.R., 2005. Role of CYP2E1 in the epoxidation of acrylamide to glycidamide and formation of DNA and hemoglobin adducts. *Toxicological Sciences*, *88*(2), pp.311-318.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A. and Smith, H.O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), pp.343-345.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A. and Lim, W.A., 2013. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, *154*(2), pp.442-451.

Granvogl, M. and Schieberle, P., 2006. Thermally generated 3-aminopropionamide as a transient intermediate in the formation of acrylamide. *Journal of Agricultural and Food Chemistry*, *54*(16), pp.5933-5938.

Granvogl, M., Wieser, H., Koehler, P., Von Tucher, S. and Schieberle, P., 2007. Influence of sulfur fertilization on the amounts of free amino acids in wheat. Correlation with baking properties as well as with 3-aminopropionamide and acrylamide generation during baking. *Journal of Agricultural and Food Chemistry*, *55*(10), pp.4271-4277.

Gratz, S.J., Ukken, F.P., Rubinstein, C.D., Thiede, G., Donohue, L.K., Cummings, A.M. and O'Connor-Giles, K.M., 2014. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. *Genetics*, *196*(4), pp.961-971.

Grohmann, L., Keilwagen, J., Duensing, N., Dagand, E., Hartung, F., Wilhelm, R., Bendiek, J. and Sprink, T., 2019. Detection and identification of genome editing in plants: challenges and opportunities. *Frontiers in Plant Science*, *10*, p.236.

Guilinger, J.P., Thompson, D.B. and Liu, D.R., 2014. Fusion of catalytically inactive Cas9 to *Fok*I nuclease improves the specificity of genome modification. *Nature Biotechnology*, *32*(6), p.577.

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J. and Joly, J.S., 2016. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biology*, *17*(1), p.148.

Hahn, F., Eisenhut, M., Mantegazza, O. and Weber, A.P., 2018. Homology-directed repair of a defective glabrous gene in Arabidopsis with Cas9-based gene targeting. *Frontiers in Plant Science*, *9*, p.424.

Halford, N.G., Curtis, T.Y., Muttucumaru, N., Postles, J. and Mottram, D.S., 2011. Sugars in crop plants. *Annals of Applied Biology*, *158*(1), pp.1-25.

Halford, N.G., Muttucumaru, N., Powers, S.J., Gillatt, P.N., Hartley, L., Elmore, J.S. and Mottram, D.S., 2012. Concentrations of free amino acids and sugars in nine potato varieties: effects of storage and relationship with acrylamide formation. *Journal of Agricultural and Food Chemistry*, *60*(48), pp.12044-12055.

Hedegaard, R.V., Frandsen, H. and Skibsted, L.H., 2008. Kinetics of formation of acrylamide and Schiff base intermediates from asparagine and glucose. *Food Chemistry*, *108*(3), pp.917-925.

Heigwer, F., Kerr, G. and Boutros, M., 2014. E-CRISP: fast CRISPR target site identification. *Nature Methods*, *11*(2), pp.122-123.

Herrera-Estrella, L., Depicker, A., Van Montagu, M. and Schell, J., 1983. Expression of chimaeric genes transferred into plant cells using a Ti-plasmid-derived vector. *Nature*, *303*(5914), pp.209-213.

Hey, S.J., Byrne, E. and Halford, N.G., 2009. The interface between metabolic and stress signalling. *Annals of Botany*, *105*(2), pp.197-203.

Hickey, J.M., Chiurugwi, T., Mackay, I., Powell, W., Eggen, A., Kilian, A., Jones, C., Canales, C., Grattapaglia, D., Bassi, F. and Atlin, G., 2017. Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nature Genetics*, *49*(9), p.1297.

Hiei, Y. and Komari, T., 2008. Agrobacterium-mediated transformation of rice using immature embryos or calli induced from mature seed. *Nature Protocols*, *3*(5), pp.824-834.

Hiei, Y., Ishida, Y. and Komari, T., 2014. Progress of cereal transformation technology mediated by *Agrobacterium tumefaciens*. *Frontiers in Plant Science*, *5*, p.628.

Hinnebusch, A.G., 1992. General and pathway-specific regulatory mechanisms controlling the synthesis of amino acid/biosynthetic enzymes in *Saccharomyces cerevisae*. *The molecular and Cellular Biology of the Yeast Saccharomyces: Gene Expression*, pp.319-414.

Hodge, J.E., 1953. Dehydrated foods, chemistry of browning reactions in model systems. *Journal of Agricultural and Food Chemistry*, *1*(15), pp.928-943.

Hogervorst, J.G., Schouten, L.J., Konings, E.J., Goldbohm, R.A. and van den Brandt, P.A., 2008. Dietary acrylamide intake and the risk of renal cell, bladder, and prostate cancer. *The American Journal of Clinical Nutrition*, *87*(5), pp.1428-1438.

Hogervorst, J.G., Schouten, L.J., Konings, E.J., Goldbohm, R.A. and van den Brandt, P.A., 2007. A prospective study of dietary acrylamide intake and the risk of endometrial, ovarian, and breast cancer. *Cancer Epidemiology and Prevention Biomarkers*, *16*(11), pp.2304-2313.

Holkers, M., Maggio, I., Liu, J., Janssen, J.M., Miselli, F., Mussolino, C., Recchia, A., Cathomen, T. and Goncalves, M.A., 2013. Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Research*, *41*(5), pp.e63-e63.

Howells, R.M., Craze, M., Bowden, S. and Wallington, E.J., 2018. Efficient generation of stable, heritable gene edits in wheat using CRISPR/Cas9. *BMC Plant biology*, *18*(1), p.215.

Hough, S.H., Ajetunmobi, A., Brody, L., Humphryes-Kirilov, N. and Perello, E., 2016. Desktop genetics.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O. and Cradick, T.J., 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, *31*(9), pp.827-832.

Huang, D.W., Sherman, B.T. and Lempicki, R.A., 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, *37*(1), pp.1-13.

Hummel, M., Rahmani, F., Smeekens, S. and Hanson, J., 2009. Sucrose-mediated translational control. *Annals of Botany*, *104*(1), pp.1-7.

Igarashi, D., Ishizaki, T., Totsuka, K. and Ohsumi, C., 2009. ASN2 is a key enzyme in asparagine biosynthesis under ammonium sufficient conditions. *Plant Biotechnology*, *26*(1), pp.153-159.

International Agency for Research on Cancer (1994) IARC Monographs on the Evaluation of Carcinogenic Risks to Humans Vol 60. Some Industrial Chemicals. International Agency for Research on Cancer (IARC), Lyon.

International Brachypodium Initiative, 2010. Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature*, *463*(7282), p.763.

International Wheat Genome Sequencing Consortium, 2014. A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science*, *345*(6194).

Iorizzo, M., Ellison, S., Senalik, D., Zeng, P., Satapoomin, P., Huang, J., Bowman, M., Iovene, M., Sanseverino, W., Cavagnaro, P. and Yildiz, M., 2016. A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nature Genetics*, *48*(6), pp.657-666.

Ishida, Y., Tsunashima, M., Hiei, Y. and Komari, T., 2015. Wheat (Triticum aestivum L.) transformation using immature embryos. In Agrobacterium Protocols (pp. 189-198). Springer, New York, NY.

Iyer, L.M., Kumpatla, S.P., Chandrasekharan, M.B. and Hall, T.C., 2000. Transgene silencing in monocots. *Plant Molecular Biology*, *43*(2-3), pp.323-346.

Jacquemin, J., Bhatia, D., Singh, K. and Wing, R.A., 2013. The International Oryza Map Alignment Project: development of a genus-wide comparative genomics platform to help solve the 9 billion-people question. *Current Opinion in Plant Biology*, *16*(2), pp.147-156.

Joint, F.A.O., WHO Expert Committee on Food Additives and World Health Organization, 2006. Evaluation of certain food contaminants: sixty-fourth report of the Joint FAO/WHO Expert Committee on Food Additives. World Health Organization.

Joint, F.A.O., World Health Organization and WHO Expert Committee on Food Additives, 2011. Evaluation of certain contaminants in food: seventy-second [72nd] report of the Joint FAO/WHO Expert Committee on Food Additives. World Health Organization.

Jefferies, R.A. and MacKerron, D.K.L., 1987. Observations on the incidence of tuber growth cracking in relation to weather patterns. *Potato Research*, 30(4), pp.613-623.

Jia, H. and Wang, N., 2014. Targeted genome editing of sweet orange using Cas9/sgRNA. *PloS One*, *9*(4), p.e93806.

Jia, H., Xu, J., Orbović, V., Zhang, Y. and Wang, N., 2017. Editing citrus genome via SaCas9/sgRNA system. *Frontiers in Plant Science*, *8*, p.2135.

Jiang, F., Zhou, K., Ma, L., Gressel, S. and Doudna, J.A., 2015. A Cas9–guide RNA complex preorganized for target DNA recognition. *Science*, *348*(6242), pp.1477-1481.

Jiang, W., Bikard, D., Cox, D., Zhang, F. and Marraffini, L.A., 2013a. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology*, *31*(3), pp.233-239.

Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B. and Weeks, D.P., 2013b. Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Research*, *41*(20), pp.e188-e188.

Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M.C., Wang, B., Campbell, M.S., Stein, J.C., Wei, X., Chin, C.S. and Guill, K., 2017. Improved maize reference genome with single-molecule technologies. *Nature*, *546*(7659), pp.524-527.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E., 2012. A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), pp.816-821.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E. and Doudna, J., 2013. RNA-programmed genome editing in human cells. *elife*, *2*, p.e00471.

Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S. and Kaplan, M., 2014. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science*, *343*(6176).

Jones, H.D., Doherty, A. and Wu, H., 2005. Review of methodologies and a protocol for the Agrobacterium-mediated transformation of wheat. *Plant Methods*, *1*(1), p.5.

Jukes, T.H. and Cantor, C.R., 1969. Evolution of protein molecules. *Mammalian Protein Metabolism*, *3*, pp.21-132.

Karley, A.J., Douglas, A.E. and Parker, W.E., 2002. Amino acid composition and nutritional quality of potato leaf phloem sap for aphids. *Journal of Experimental Biology*, *205*(19), pp.3009-3018.

Kartha, K.K., Chibbar, R., Georges, F., Leung, N., Caswell, K., Kendall, E. and Qureshi, J., 1989. Transient expression of chloramphenicol acetyltransferase (CAT) gene in barley cell cultures and immature embryos through microprojectile bombardment. *Plant Cell Reports*, *8*(8), pp.429-432.

Kausch, A.P., Adams, T.R., Mangano, M., Zachwieja, S.J., Gordon-Kamm, W., Daines, R., Willetts, N.G., Chambers, S.A., Adams, W., Anderson, A. and Williams, G., 1995. Effects of microprojectile bombardment on embryogenic suspension cell cultures of maize (*Zea mays* L.) used for genetic transformation. *Planta*, *196*(3), pp.501-509.

Kawachi, T., Sueyoshi, K., Nakajima, A., Yamagata, H., Sugimoto, T. and Oji, Y., 2002. Expression of asparagine synthetase in rice (*Oryza sativa*) roots in response to nitrogen. *Physiologia Plantarum*, *114*(1), pp.41-46.

Kawahara, Y., de la Bastide, M., Hamilton, J.P., Kanamori, H., McCombie, W.R., Ouyang, S., Schwartz, D.C., Tanaka, T., Wu, J., Zhou, S. and Childs, K.L., 2013. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice*, *6*(1), p.4.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C. and Thierer, T., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), pp.1647-1649.

Kim, D., Alptekin, B. and Budak, H., 2018. CRISPR/Cas9 genome editing in wheat. *Functional & Integrative Genomics*, *18*(1), pp.31-41.

Kim, H., Kim, S.T., Ryu, J., Choi, M.K., Kweon, J., Kang, B.C., Ahn, H.M., Bae, S., Kim, J., Kim, J.S. and Kim, S.G., 2016. A simple, flexible and high-throughput cloning system for plant genome editing via CRISPR-Cas system. *Journal of Integrative Plant Biology*, *58*(8), pp.705-712.

Kim, H., Kim, S.T., Ryu, J., Kang, B.C., Kim, J.S. and Kim, S.G., 2017. CRISPR/Cpf1-mediated DNA-free plant genome editing. *Nature Communications*, 8(1), pp.1-7.

Kim, Y.G., Cha, J. and Chandrasegaran, S., 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences*, *93*(3), pp.1156-1160.

Klein, T.M., Gradziel, T., Fromm, M.E. and Sanford, J.C., 1988. Factors influencing gene delivery into *Zea mays* cells by high-velocity microprojectiles. *Biotechnology*, *6*(5), pp.559-563.

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z. and Joung, J.K., 2016. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, *529*(7587), pp.490-495.

Koch, W., Kwart, M., Laubner, M., Heineke, D., Stransky, H., Frommer, W.B. and Tegeder, M., 2003. Reduced amino acid content in transgenic potato tubers due to antisense inhibition of the leaf H+/amino acid symporter StAAP1. *The Plant Journal*, *33*(2), pp.211-220.

Kohli, A., Twyman, R.M., Abranches, R., Wegel, E., Stoger, E. and Christou, P., 2003. Transgene integration, organization and interaction in plants. *Plant Molecular Biology*, *52*(2), pp.247-258.

Koren, S., Schatz, M.C., Walenz, B.P., Martin, J., Howard, J.T., Ganapathy, G., Wang, Z., Rasko, D.A., McCombie, W.R., Jarvis, E.D. and Phillippy, A.M., 2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nature Biotechnology*, *30*(7), pp.693-700.

Krasileva, K.V., Vasquez-Gross, H.A., Howell, T., Bailey, P., Paraiso, F., Clissold, L., Simmonds, J., Ramirez-Gonzalez, R.H., Wang, X., Borrill, P. and Fosker, C., 2017. Uncovering hidden variation in polyploid wheat. *Proceedings of the National Academy of Sciences*, *114*(6), pp.913-921.

Krause, G. and Weis, E., 1991. Chlorophyll fluorescence and photosynthesis: the basics. *Annual Review of Plant Biology*, *42*(1), pp.313-349.

Kuromori, T., Wada, T., Kamiya, A., Yuguchi, M., Yokouchi, T., Imura, Y., Takabe, H., Sakurai, T., Akiyama, K., Hirayama, T. and Okada, K., 2006. A trial of phenome analysis using 4000 Ds-

insertional mutants in gene-coding regions of Arabidopsis. *The Plant Journal*, *47*(4), pp.640-651.

Kuscu, C., Arslan, S., Singh, R., Thorpe, J. and Adli, M., 2014. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nature Biotechnology*, *32*(7), pp.677-683.

Lageix, S., Lanet, E., Pouch-Pélissier, M.N., Espagnol, M.C., Robaglia, C., Deragon, J.M. and Pélissier, T., 2008. Arabidopsis eIF2α kinase GCN2 is essential for growth in stress conditions and is activated by wounding. *BMC Plant Biology*, *8*(1), pp.1-9.

Lam, E.T., Hastie, A., Lin, C., Ehrlich, D., Das, S.K., Austin, M.D., Deshpande, P., Cao, H., Nagarajan, N., Xiao, M. and Kwok, P.Y., 2012. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nature Biotechnology*, *30*(8), pp.771-776.

Lam, H.M., Hsieh, M.H. and Coruzzi, G., 1998. Reciprocal regulation of distinct asparagine synthetase genes by light and metabolites in Arabidopsis thaliana. *The Plant Journal*, *16*(3), pp.345-353.

Lam, H.M., Peng, S.S. and Coruzzi, G.M., 1994. Metabolic regulation of the gene encoding glutamine-dependent asparagine synthetase in Arabidopsis thaliana. *Plant Physiology*, *106*(4), pp.1347-1357.

Lam, H.M., Wong, P., Chan, H.K., Yam, K.M., Chen, L., Chow, C.M. and Coruzzi, G.M., 2003. Overexpression of the ASN1 gene enhances nitrogen status in seeds of Arabidopsis. *Plant Physiology*, *132*(2), pp.926-935.

Langmead, B. and Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), p.357.

Lea, P.J., Sodek, L., Parry, M.A., Shewry, P.R. and Halford, N.G., 2007. Asparagine in plants. *Annals of Applied Biology*, *150*(1), pp.1-26.

Li, H. and Homer, N., 2010. A survey of sequence alignment algorithms for next-generation sequencing. *Briefings in Bioinformatics*, *11*(5), pp.473-483.

Li, J.F., Norville, J.E., Aach, J., McCormack, M., Zhang, D., Bush, J., Church, G.M. and Sheen, J., 2013. Multiplex and homologous recombination–mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. *Nature Biotechnology*, *31*(8), pp.688-691.

Li, M., Schönberg, A., Schaefer, M., Schroeder, R., Nasidze, I. and Stoneking, M., 2010. Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. *The American Journal of Human Genetics*, *87*(2), pp.237-249.

Li, Q., Guo, L., Wang, H., Zhang, Y., Fan, C. and Shen, Y., 2019. In silico genome-wide identification and comprehensive characterization of the *BES1* gene family in soybean. *Heliyon*, *5*(6), p.e01868.

Li, T., Huang, S., Jiang, W.Z., Wright, D., Spalding, M.H., Weeks, D.P. and Yang, B., 2011. TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and *Fok*I DNA-cleavage domain. *Nucleic Acids Research*, *39*(1), pp.359-372.

Li, Z., Liu, Z.B., Xing, A., Moon, B.P., Koellhoffer, J.P., Huang, L., Ward, R.T., Clifton, E., Falco, S.C. and Cigan, A.M., 2015. Cas9-guide RNA directed genome editing in soybean. *Plant Physiology*, *169*(2), pp.960-970.

Liang, Z., Chen, K., Li, T., Zhang, Y., Wang, Y., Zhao, Q., Liu, J., Zhang, H., Liu, C., Ran, Y. and Gao, C., 2017. Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. *Nature Communications*, *8*(1), pp.1-5.

Liang, Z., Chen, K., Zhang, Y., Liu, J., Yin, K., Qiu, J.L. and Gao, C., 2018. Genome editing of bread wheat using biolistic delivery of CRISPR/Cas9 in vitro transcripts or ribonucleoproteins. *Nature Protocols*, *13*(3), p.413.

Liang, Z., Zhang, K., Chen, K. and Gao, C., 2014. Targeted mutagenesis in Zea mays using TALENs and the CRISPR/Cas system. *Journal of Genetics and Genomics*, *41*(2), pp.63-68.

Lieber, M.R., 2010. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual Review of Biochemistry*, *79*, pp.181-211.

Lieberman-Aiden, E., Van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O. and Sandstrom, R., 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, *326*(5950), pp.289-293.

Lin, Y., Cradick, T.J., Brown, M.T., Deshmukh, H., Ranjan, P., Sarode, N., Wile, B.M., Vertino, P.M., Stewart, F.J. and Bao, G., 2014. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Research*, *42*(11), pp.7473-7485.

Ling, H.Q., Zhao, S., Liu, D., Wang, J., Sun, H., Zhang, C., Fan, H., Li, D., Dong, L., Tao, Y. and Gao, C., 2013. Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature*, *496*(7443), pp.87-90.

Liu, H., Wei, Z., Dominguez, A., Li, Y., Wang, X. and Qi, L.S., 2015a. CRISPR-ERA: a comprehensive design tool for CRISPR-mediated gene editing, repression and activation. *Bioinformatics*, *31*(22), pp.3676-3678.

Liu, W., Xie, X., Ma, X., Li, J., Chen, J. and Liu, Y.G., 2015b. DSDecode: a web-based tool for decoding of sequencing chromatograms for genotyping of targeted mutations. *Molecular Plant*, *8*(9), pp.1431-1433.

Liu, Y., Schiff, M. and Dinesh-Kumar, S.P., 2002. Virus-induced gene silencing in tomato. *The Plant Journal*, *31*(6), pp.777-786.

Liu, Y., Tao, W., Wen, S., Li, Z., Yang, A., Deng, Z. and Sun, Y., 2015. In vitro CRISPR/Cas9 system for efficient targeted DNA editing. *MBio*, *6*(6).

Lloyd, A., Plaisier, C.L., Carroll, D. and Drews, G.N., 2005. Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. *Proceedings of the National Academy of Sciences*, *102*(6), pp.2232-2237.

Lorenz, R., Bernhart, S.H., Zu Siederdissen, C.H., Tafer, H., Flamm, C., Stadler, P.F. and Hofacker, I.L., 2011. ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, 6(1), p.26.

Lowder, L.G., Zhang, D., Baltes, N.J., Paul, J.W., Tang, X., Zheng, X., Voytas, D.F., Hsieh, T.F., Zhang, Y. and Qi, Y., 2015. A CRISPR/Cas9 toolbox for multiplexed plant genome editing and transcriptional regulation. *Plant Physiology*, *169*(2), pp.971-985.

Lu, Y. and Zhu, J.K., 2017. Precise editing of a target base in the rice genome using a modified CRISPR/Cas9 system. *Molecular Plant*, *10*(3), pp.523-525.

Luo, L., Pan, S., Liu, X., Wang, H. and Xu, G., 2017a. Nitrogen deficiency inhibits cell division-determined elongation, but not initiation, of rice tiller buds. *Israel Journal of Plant Sciences*, *64*(3-4), pp.32-40.

Luo, L., Qin, R., Liu, T., Yu, M., Yang, T. and Xu, G., 2019. *OsASN1* plays a critical role in asparagine-dependent rice development. *International Journal of Molecular Sciences*, *20*(1), p.130.

Luo, M.C., Gu, Y.Q., Puiu, D., Wang, H., Twardziok, S.O., Deal, K.R., Huo, N., Zhu, T., Wang, L., Wang, Y. and McGuire, P.E., 2017b. Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature*, *551*(7681), pp.498-502.

Ye, A.Y., Zheng, W. and Kong, L., 2013. A guide RNA sequence design platform for the CRISPR/Cas9 system for model organism genomes. *BioMed Research International*, *2013*.

Ma, X., Chen, L., Zhu, Q., Chen, Y. and Liu, Y.G., 2015a. Rapid decoding of sequence-specific nuclease-induced heterozygous and biallelic mutations by direct sequencing of PCR products. *Molecular Plant*, *8*(8), pp.1285-1287.

Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R., Wang, B., Yang, Z., Li, H., Lin, Y. and Xie, Y., 2015b. A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *Molecular Plant*, *8*(8), pp.1274-1284.

Ma, X., Zhu, Q., Chen, Y. and Liu, Y.G., 2016. CRISPR/Cas9 platforms for genome editing in plants: developments and applications. *Molecular Plant*, *9*(7), pp.961-974.

Maccaferri, M., Harris, N.S., Twardziok, S.O., Pasam, R.K., Gundlach, H., Spannagl, M., Ormanbekova, D., Lux, T., Prade, V.M., Milner, S.G. and Himmelbach, A., 2019. Durum wheat genome highlights past domestication signatures and future improvement targets. *Nature Genetics*, *51*(5), pp.885-895.

Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H. and Joung, J.K., 2013. CRISPR RNA–guided activation of endogenous human genes. *Nature Methods*, *10*(10), pp.977-979.

Maillard, L.C. (1912) Action des acides aminés sur les sucres: formation des mélanoïdines par voie méthodique. *Compte-rendu de l'Académie des Sciences*, *154*, 66-68.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F. and Van Der Oost, J., 2011. Evolution and classification of the CRISPR–Cas systems. *Nature Reviews Microbiology*, *9*(6), pp.467-477.

Makita, Y., Shimada, S., Kawashima, M., Kondou-Kuriyama, T., Toyoda, T. and Matsui, M., 2015. MOROKOSHI: transcriptome database in *Sorghum bicolor*. *Plant and Cell Physiology*, *56*(1).

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L. and Church, G.M., 2013a. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology*, *31*(9), pp.833-838.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. and Church, G.M., 2013b. RNA-guided human genome engineering via Cas9. *Science*, *339*(6121), pp.823-826.

Mao, Y., Zhang, H., Xu, N., Zhang, B., Gou, F. and Zhu, J.K., 2013. Application of the CRISPR–Cas system for efficient genome engineering in plants. *Molecular Plant*, *6*(6), pp.2008-2011.

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I. and Lanczycki, C.J., 2015. CDD: NCBI's conserved domain database. *Nucleic Acids Research*, *43*(D1), pp.222-226.

Marcussen, T., Sandve, S.R., Heier, L., Spannagl, M., Pfeifer, M., Jakobsen, K.S., Wulff, B.B., Steuernagel, B., Mayer, K.F., Olsen, O.A. and International Wheat Genome Sequencing Consortium, 2014. Ancient hybridizations among the ancestral genomes of bread wheat. *Science*, *345*(6194).

Martinek, P., Klem, K., Vanova, M., Bartackova, V., Vecerkova, L., Bucher, P. and Hajslova, J., 2009. Effects of nitrogen nutrition, fungicide treatment and wheat genotype on free

asparagine and reducing sugars content as precursors of acrylamide formation in bread. *Plant Soil and Environment*, *55*(5), pp.187-195.

Martins, S.I., Jongen, W.M. and Van Boekel, M.A., 2000. A review of Maillard reaction in food and implications to kinetic modelling. *Trends in Food Science & Technology*, *11*(9-10), pp.364-373.

Mascher, M., Muehlbauer, G.J., Rokhsar, D.S., Chapman, J., Schmutz, J., Barry, K., Muñoz-Amatriaín, M., Close, T.J., Wise, R.P., Schulman, A.H. and Himmelbach, A., 2013. Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *The Plant Journal*, *76*(4), pp.718-727.

Masclaux-Daubresse, C., Reisdorf-Cren, M. and Orsel, M., 2008. Leaf nitrogen remobilisation for plant development and grain filling. *Plant Biology*, *10*, pp.23-36.

Masella, A.P., Bartram, A.K., Truszkowski, J.M., Brown, D.G. and Neufeld, J.D., 2012. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics*, *13*(1), p.31.

Matsuoka, Y., 2011. Evolution of polyploid Triticum wheats under cultivation: the role of domestication, natural hybridization and allopolyploid speciation in their diversification. *Plant and Cell Physiology*, *52*(5), pp.750-764.

Mayer, K.F., Taudien, S., Martis, M., Šimková, H., Suchánková, P., Gundlach, H., Wicker, T., Petzold, A., Felder, M., Steuernagel, B. and Scholz, U., 2009. Gene content and virtual gene order of barley chromosome 1H. *Plant Physiology*, *151*(2), pp.496-505.

Mayer, K.F., Waugh, R., Brown, J.W., Schulman, A., Langridge, P., Platzer, M., Fincher, G.B., Muehlbauer, G.J., Sato, K., Close, T.J. and Wise, R.P., 2012. International Barley Genome Sequencing Consortium. *Nature*, *491*, p.711.

McCabe, D.E. and Martinell, B.J., 1993. Transformation of elite cotton cultivars via particle bombardment of meristems. *Bio/technology*, *11*(5), pp.596-598.

McCormick, R.F., Truong, S.K., Sreedasyam, A., Jenkins, J., Shu, S., Sims, D., Kennedy, M., Amirebrahimi, M., Weers, B.D., McKinley, B. and Mattison, A., 2018. The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *The Plant Journal*, *93*(2), pp.338-354.

McFadden, E.S. and Sears, E.R., 1946. The origin of Triticum spelta and its free-threshing hexaploid relatives. *Journal of Heredity*, *37*(3), pp.81-89.

Menz, J., Modrzejewski, D., Hartung, F., Wilhelm, R. and Sprink, T., 2020. Genome edited crops touch the market: a view on the global development and regulatory environment. *Frontiers in Plant Science*, *11*.

Miao, J., Guo, D., Zhang, J., Huang, Q., Qin, G., Zhang, X., Wan, J., Gu, H. and Qu, L.J., 2013. Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Research*, *23*(10), pp.1233-1236.

Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H.Y., El-Gebali, S., Fraser, M.I. and Gough, J., 2019. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Research*, *47*(D1), pp.351-360.

Mol, J.N.M., Van der Krol, A.R., Van Tunen, A.J., Van Blokland, R., de Lange, P. and Stuitje, A.R., 1990. Regulation of plant gene expression by antisense RNA. *FEBS letters*, *268*(2), pp.427-430.

Møller, M.G., Taylor, C., Rasmussen, S.K. and Holm, P.B., 2003. Molecular cloning and characterisation of two genes encoding asparagine synthetase in barley (*Hordeum vulgare*

L.). *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, *1628*(2), pp.123-132.

Montague, T.G., Cruz, J.M., Gagnon, J.A., Church, G.M. and Valen, E., 2014. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Research*, *42*(W1), pp.401-407.

Morikawa, H., Iida, A. and Yamada, Y., 1989. Transient expression of foreign genes in plant cells and tissues obtained by a simple biolistic device (particle-gun). *Applied Microbiology and Biotechnology*, *31*(3), pp.320-322.

Moscou, M.J. and Bogdanove, A.J., 2009. A simple cipher governs DNA recognition by TAL effectors. *Science*, *326*(5959), pp.1501-1501.

Mottram, D.S., 2007. The Maillard reaction: source of flavour in thermally processed foods. In *Flavours and Fragrances* (pp. 269-283). Springer, Berlin, Heidelberg.

Mottram, D.S., Wedzicha, B.L. and Dodson, A.T., 2002. Acrylamide is formed in the Maillard reaction. *Nature*, *419*(6906), pp.448-449.

Muttucumaru, N., Halford, N.G., Elmore, J.S., Dodson, A.T., Parry, M., Shewry, P.R. and Mottram, D.S., 2006. Formation of high levels of acrylamide during the processing of flour derived from sulfate-deprived wheat. *Journal of Agricultural and Food Chemistry*, *54*(23), pp.8951-8955.

Muttucumaru, N., Keys, A.J., Parry, M.A., Powers, S.J. and Halford, N.G., 2014. Photosynthetic assimilation of ^{14}C into amino acids in potato (*Solanum tuberosum*) and asparagine in the tubers. *Planta*, *239*(1), pp.161-170.

Nakajima, I., Ban, Y., Azuma, A., Onoue, N., Moriguchi, T., Yamamoto, T., Toki, S. and Endo, M., 2017. CRISPR/Cas9-mediated targeted mutagenesis in grape. *PLoS One*, *12*(5), p.e0177966.

Nakano, K., Suzuki, T., Hayakawa, T. and Yamaya, T., 2000. Organ and cellular localization of asparagine synthetase in rice plants. *Plant and Cell Physiology*, *41*(7), pp.874-880.

Nam, J., Matthysse, A.G. and Gelvin, S.B., 1997. Differences in susceptibility of Arabidopsis ecotypes to crown gall disease may result from a deficiency in T-DNA integration. *The Plant Cell*, *9*(3), pp.317-333.

Navrotskyi, S., Baenziger, P.S., Regassa, T., Guttieri, M.J. and Rose, D.J., 2018. Variation in asparagine concentration in Nebraska wheat. *Cereal Chemistry*, *95*(2), pp.264-273.

Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J.D. and Kamoun, S., 2013. Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. *Nature Biotechnology*, *31*(8), pp.691-693.

Nekrasov, V., Wang, C., Win, J., Lanz, C., Weigel, D. and Kamoun, S., 2017. Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. *Scientific Reports*, *7*(1), pp.1-6.

Nishijima, R., Yoshida, K., Motoi, Y., Sato, K. and Takumi, S., 2016. Genome-wide identification of novel genetic markers from RNA sequencing assembly of diverse *Aegilops tauschii* accessions. *Molecular Genetics and Genomics*, *291*(4), pp.1681-1694.

Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F. and Nureki, O., 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, *156*(5), pp.935-949.

Noctor, G. and Foyer, C.H., 1998. Simultaneous measurement of foliar glutathione, γ-glutamylcysteine, and amino acids by high-performance liquid chromatography: comparison with two other assay methods for glutathione. *Analytical Biochemistry*, *264*(1), pp.98-110.

O'Geen, H., Henry, I.M., Bhakta, M.S., Meckler, J.F. and Segal, D.J., 2015. A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. *Nucleic Acids Research*, *43*(6), pp.3389-3404.

Oard, J.H., Paige, D.F., Simmonds, J.A. and Gradziel, T.M., 1990. Transient gene expression in maize, rice, and wheat cells using an airgun apparatus. *Plant Physiology*, *92*(2), pp.334-339.

Ohashi, M., Ishiyama, K., Kojima, S., Konishi, N., Nakano, K., Kanno, K., Hayakawa, T. and Yamaya, T., 2015. Asparagine synthetase1, but not asparagine synthetase2, is responsible for the biosynthesis of asparagine following the supply of ammonium to rice roots. *Plant and Cell Physiology*, *56*(4), pp.769-778.

Okada, A., Arndell, T., Borisjuk, N., Sharma, N., Watson-Haigh, N.S., Tucker, E.J., Baumann, U., Langridge, P. and Whitford, R., 2019. CRISPR/Cas9-mediated knockout of *Ms1* enables the rapid generation of male-sterile hexaploid wheat lines for use in hybrid seed production. *Plant Biotechnology Journa*l, *17*(10), pp.1905-1913.

Oliva, R., Ji, C., Atienza-Grande, G., Huguet-Tapia, J.C., Perez-Quintero, A., Li, T., Eom, J.S., Li, C., Nguyen, H., Liu, B. and Auguy, F., 2019. Broad-spectrum resistance to bacterial blight in rice using genome editing. *Nature Biotechnology*, *37*(11), pp.1344-1350.

Oliveira, I.C., Brenner, E., Chiu, J., Hsieh, M.H., Kouranov, A., Lam, H.M., Shin, M.J. and Coruzzi, G., 2001. Metabolite and light regulation of metabolism in plants: lessons from the study of a single biochemical pathway. *Brazilian Journal of Medical and Biological Research*, *34*(5), pp.567-575.

Osakabe, K., Osakabe, Y. and Toki, S., 2010. Site-directed mutagenesis in Arabidopsis using custom-designed zinc finger nucleases. *Proceedings of the National Academy of Sciences*, *107*(26), pp.12034-12039.

Pang, J., Fu, J., Zong, N., Wang, J., Song, D., Zhang, X., He, C., Fang, T., Zhang, H., Fan, Y. and Wang, G., 2019. Kernel size-related genes revealed by an integrated eQTL analysis during early maize kernel development. *The Plant Journal*, *98*(1), pp.19-32.

Park, J., Bae, S. & J.S, K., 2015. Cas-Designer: A web-based tool for choice of CRISPR-Cas9 target sites. *Bioinformatics*, *31*, pp.4014-4016.

Partier, A., Gay, G., Tassy, C., Beckert, M., Feuillet, C. and Barret, P., 2017. Molecular and FISH analyses of a 53-kbp intact DNA fragment inserted by biolistics in wheat (*Triticum aestivum* L.) genome. *Plant Cell Reports*, *36*(10), pp.1547-1559.

Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A. and Liu, D.R., 2013. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnology*, *31*(9), pp.839-843.

Pearce, S., Saville, R., Vaughan, S.P., Chandler, P.M., Wilhelm, E.P., Sparks, C.A., Al-Kaff, N., Korolev, A., Boulton, M.I., Phillips, A.L. and Hedden, P., 2011. Molecular characterization of *Rht-1* dwarfing genes in hexaploid wheat. *Plant Physiology*, *157*(4), pp.1820-1831.

Pelucchi, C., Bosetti, C., Galeone, C. and La Vecchia, C., 2015. Dietary acrylamide and cancer risk: an updated meta-analysis. *International Journal of Cancer*, *136*(12), pp.2912-2922.

Peng, J.H., Sun, D. and Nevo, E., 2011. Domestication evolution, genetics and genomics in wheat. *Molecular Breeding*, *28*(3), p.281.

Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W. and Guilak, F., 2013. RNA-guided gene activation by CRISPR-Cas9–based transcription factors. *Nature Methods*, *10*(10), pp.973-976.

Postles, J., Curtis, T.Y., Powers, S.J., Elmore, J.S., Mottram, D.S. and Halford, N.G., 2016. Changes in free amino acid concentration in rye grain in response to nitrogen and sulfur availability, and expression analysis of genes involved in asparagine metabolism. *Frontiers in Plant Science*, *7*, p.917.

Postles, J., Powers, S.J., Elmore, J.S., Mottram, D.S. and Halford, N.G., 2013. Effects of variety and nutrient availability on the acrylamide-forming potential of rye grain. *Journal of Cereal Science*, *57*(3), pp.463-470.

Powers, S.J., Mottram, D.S., Curtis, A. and Halford, N.G., 2013. Acrylamide concentrations in potato crisps in Europe from 2002 to 2011. *Food Additives & Contaminants: Part A*, *30*(9), pp.1493-1500.

Powers, S.J., Mottram, D.S., Curtis, A. and Halford, N.G., 2017. Acrylamide levels in potato crisps in Europe from 2002 to 2016. *Food Additives & Contaminants: Part A*, *34*(12), pp.2085-2100.

Puchta, H. and Fauser, F., 2014. Synthetic nucleases for genome engineering in plants: prospects for a bright future. *The Plant Journal*, *78*(5), pp.727-741.

Puchta, H., 2005. The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *Journal of Experimental Botany*, *56*(409), pp.1-14.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P. and Lim, W.A., 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, *152*(5), pp.1173-1183.

Raffan, S. and Halford, N.G., 2020. Cereal asparagine synthetase genes. *Annals of Applied Biology*.

Ramirez, C.L., Foley, J.E., Wright, D.A., Müller-Lerch, F., Rahman, S.H., Cornu, T.I., Winfrey, R.J., Sander, J.D., Fu, F., Townsend, J.A. and Cathomen, T., 2008. Unexpected failure rates for modular assembly of engineered zinc fingers. *Nature Methods*, *5*(5), pp.374-375.

Ramírez-González, R.H., Borrill, P., Lang, D., Harrington, S.A., Brinton, J., Venturini, L., Davey, M., Jacobs, J., Van Ex, F., Pasha, A. and Khedikar, Y., 2018. The transcriptional landscape of polyploid wheat. *Science*, *361*(6403).

Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S. and Koonin, E.V., 2015. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature*, *520*(7546), pp.186-191.

Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y. and Zhang, F., 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, *154*(6), pp.1380-1389.

Rapp, M., Schwadorf, K., Leiser, W.L., Würschum, T. and Longin, C.F.H., 2018. Assessing the variation and genetic architecture of asparagine content in wheat: What can plant breeding contribute to a reduction in the acrylamide precursor?. *Theoretical and Applied Genetics*, *131*(11), pp.2427-2437.

Reggiardo, M.I., Arana, J.L., Orsaria, L.M., Permingeat, H.R., Spitteler, M.A. and Vallejos, R.H., 1991. Transient transformation of maize tissues by microparticle bombardment. *Plant Science*, *75*(2), pp.237-243.

Reynoso, M.A., Pauluzzi, G.C., Kajala, K., Cabanlit, S., Velasco, J., Bazin, J., Deal, R., Sinha, N.R., Brady, S.M. and Bailey-Serres, J., 2018. Nuclear transcriptomes at high resolution using retooled INTACT. *Plant Physiology*, *176*(1), pp.270-281.

Rivera, A.L., Gómez-Lim, M., Fernández, F. and Loske, A.M., 2012. Physical methods for genetic plant transformation. *Physics of Life Reviews*, *9*(3), pp.308-345.

Rodríguez-Leal, D., Lemmon, Z.H., Man, J., Bartlett, M.E. and Lippman, Z.B., 2017. Engineering quantitative trait variation for crop improvement by genome editing. *Cell*, *171*(2), pp.470-480.

Romano, A., Raemakers, K., Visser, R. and Mooibroek, H., 2001. Transformation of potato (Solanum tuberosum) using particle bombardment. *Plant Cell Reports*, *20*(3), pp.198-204.

Rommens, C.M., Yan, H., Swords, K., Richael, C. and Ye, J., 2008. Low-acrylamide French fries and potato chips. *Plant Biotechnology Journal*, *6*(8), pp.843-853.

Rosli, R., Chan, P.L., Chan, K.L., Amiruddin, N., Low, E.T.L., Singh, R., Harwood, J.L. and Murphy, D.J., 2018. In silico characterization and expression profiling of the diacylglycerol acyltransferase gene family (DGAT1, DGAT2, DGAT3 and WS/DGAT) from oil palm, Elaeis guineensis. *Plant Science*, *275*, pp.84-96.

Ruiz, M.T., Voinnet, O. and Baulcombe, D.C., 1998. Initiation and maintenance of virus-induced gene silencing. *The Plant Cell*, *10*(6), pp.937-946.

Saitou, N. and Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, *4*(4), pp.406-425.

Sakai, H., Mizuno, H., Kawahara, Y., Wakimoto, H., Ikawa, H., Kawahigashi, H., Kanamori, H., Matsumoto, T., Itoh, T. and Gaut, B.S., 2011. Retrogenes in rice (*Oryza sativa* L. ssp. *japonica*) exhibit correlated expression with their source genes. *Genome Biology and Evolution*, *3*, pp.1357-1368.

Sakai, H., Lee, S.S., Tanaka, T., Numa, H., Kim, J., Kawahara, Y., Wakimoto, H., Yang, C.C., Iwamoto, M., Abe, T. and Yamada, Y., 2013. Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics. *Plant and Cell Physiology*, *54*(2), pp.e6-e6.

Salamini, F., Özkan, H., Brandolini, A., Schäfer-Pregl, R. and Martin, W., 2002. Genetics and geography of wild cereal domestication in the near east. *Nature Reviews Genetics*, *3*(6), pp.429-441.

Salomon, S. and Puchta, H., 1998. Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. *The EMBO Journal*, *17*(20), pp.6086-6095.

Sánchez-León, S., Gil-Humanes, J., Ozuna, C.V., Giménez, M.J., Sousa, C., Voytas, D.F. and Barro, F., 2018. Low-gluten, nontransgenic wheat engineered with CRISPR/Cas9. *Plant Biotechnology Journal*, *16*(4), pp.902-910.

Sander, J.D., Dahlborg, E.J., Goodwin, M.J., Cade, L., Zhang, F., Cifuentes, D., Curtin, S.J., Blackburn, J.S., Thibodeau-Beganny, S., Qi, Y. and Pierick, C.J., 2011. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nature Methods*, *8*(1), pp.67-69.

Sanford, J.C., 1988. The biolistic process. *Trends in Biotechnology*, *6*(12), pp.299-302.

Sanford, J.C., Klein, T.M., Wolf, E.D. and Allen, N., 1987. Delivery of substances into cells and tissues using a particle bombardment process. *Particulate Science and Technology*, *5*(1), pp.27-37.

Sanford, J.C., Smith, F.D. and Russell, J.A., 1993. Optimizing the biolistic process for different biological applications. *Methods in Enzymology*, *217*, pp.483-509.

Sanjana, N.E., Cong, L., Zhou, Y., Cunniff, M.M., Feng, G. and Zhang, F., 2012. A transcription activator-like effector toolbox for genome engineering. *Nature Protocols*, *7*(1), pp.171-192.

Sasaki, T., 2005. The map-based sequence of the rice genome. *Nature*, *436*(7052), pp.793-800.

Schiml, S., Fauser, F. and Puchta, H., 2014. The CRISPR/C as system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in Arabidopsis resulting in heritable progeny. *The Plant Journal*, *80*(6), pp.1139-1150.

Schmidt, M.H.W., Vogel, A., Denton, A.K., Istace, B., Wormit, A., van de Geest, H., Bolger, M.E., Alseekh, S., Maß, J., Pfaff, C. and Schurr, U., 2017. De novo assembly of a new *Solanum pennellii* accession using nanopore sequencing. *The Plant Cell*, *29*(10), pp.2336-2348.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A. and Minx, P., 2009. The B73 maize genome: complexity, diversity, and dynamics. S*cience*, *326*(5956), pp.1112-1115.

Schulte, D., Close, T.J., Graner, A., Langridge, P., Matsumoto, T., Muehlbauer, G., Sato, K., Schulman, A.H., Waugh, R., Wise, R.P. and Stein, N., 2009. The international barley sequencing consortium—at the threshold of efficient access to the barley genome. *Plant Physiology*, *149*(1), pp.142-147.

Segerbäck, D., Calleman, C.J., Schroeder, J.L., Costa, L.G. and Faustman, E.M., 1995. Formation of N-7-(2-carbamoyl-2-hydroxyethyl) guanine in DNA of the mouse and the rat following intraperitoneal administration of [14C] acrylamide. *Carcinogenesis*, *16*(5), pp.1161-1165.

Shan, Q., Wang, Y., Li, J. and Gao, C., 2014. Genome editing in rice and wheat using the CRISPR/Cas system. *Nature Protocols*, *9*(10), pp. 2395-2410.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J.J., Qiu, J.L. and Gao, C., 2013. Targeted genome modification of crop plants using a CRISPR-Cas system. *Nature Biotechnology*, *31*(8), pp.686-688.

Shen, J., Fu, J., Ma, J., Wang, X., Gao, C., Zhuang, C., Wan, J. and Jiang, L., 2014a. Isolation, culture, and transient transformation of plant protoplasts. *Current Protocols in Cell Biology*, *63*(1), pp.2-8.

Shen, B., Zhang, W., Zhang, J., Zhou, J., Wang, J., Chen, L., Wang, L., Hodgkins, A., Iyer, V., Huang, X. and Skarnes, W.C., 2014b. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nature Methods*, *11*(4), pp.399-402.

Shepherd, L.V.T., Bradshaw, J.E., Dale, M.F.B., McNicol, J.W., Pont, S.D.A., Mottram, D.S. and Davies, H.V., 2010. Variation in acrylamide producing potential in potato: segregation of the trait in a breeding population. *Food Chemistry*, *123*(3), pp.568-573.

Shepherd, L.V.T., Pont, S.D.A., Bryan, G.J., Dale, M.F.B., Hancock, R.D., Hedley, P.E., Morris, J.A., Verrall, S.R., Hackett, C.A., McNicol, J.W. and Davies, H.V., 2013. Acrylamide forming potential of potato: predictive tools and genetic interventions. *Aspects of Applied Biology*, (116), pp.53-60.

Shewry, P.R., Franklin, J., Parmar, S., Smith, S.J. and Miflin, B.J., 1983. The effects of sulphur starvation on the amino acid and protein compositions of barley grain. *Journal of Cereal Science*, *1*(1), pp.21-31.

Shewry, P.R., Zhao, F.J., Gowa, G.B., Hawkins, N.D., Ward, J.L., Beale, M.H., Halford, N.G., Parry, M.A. and Abécassis, J., 2009. Sulphur nutrition differentially affects the distribution of asparagine in wheat grain. *Journal of Cereal Science*, *50*(3), pp.407-409.

Shrawat, A.K. and Lörz, H., 2006. Agrobacterium-mediated transformation of cereals: a promising approach crossing barriers. *Plant Biotechnology Journal*, *4*(6), pp.575-603.

Shukla, V.K., Doyon, Y., Miller, J.C., DeKelver, R.C., Moehle, E.A., Worden, S.E., Mitchell, J.C., Arnold, N.L., Gopalan, S., Meng, X. and Choi, V.M., 2009. Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature*, *459*(7245), pp.437-441.

Siebert, R. and Puchta, H., 2002. Efficient repair of genomic double-strand breaks by homologous recombination between directly repeated sequences in the plant genome. *The Plant Cell*, *14*(5), pp.1121-1131.

Sieciechowicz, K.A., Joy, K.W. and Ireland, R.J., 1988. The metabolism of asparagine in plants. *Phytochemistry*, *27*(3), pp.663-671.

Sivamani, E., DeLong, R.K. and Qu, R., 2009. Protamine-mediated DNA coating remarkably improves bombardment transformation efficiency in plant cells. *Plant Cell Reports*, *28*(2), pp.213-221.

Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X. and Zhang, F., 2016. Rationally engineered Cas9 nucleases with improved specificity. *Science*, *351*(6268), pp.84-88.

Smith, N.A., Singh, S.P., Wang, M.B., Stoutjesdijk, P.A., Green, A.G. and Waterhouse, P.M., 2000. Total silencing by intron-spliced hairpin RNAs. *Nature*, *407*(6802), pp.319-320.

Snape, J.W., Foulkes, M.J., Simmonds, J., Leverington, M., Fish, L.J., Wang, Y. and Ciavarrella, M., 2007. Dissecting gene× environmental effects on wheat yields via QTL and physiological analysis. *Euphytica*, *154*(3), pp.401-408.

Sodek, L. and Wilson, C.M., 1971. Amino acid compositions of proteins isolated from normal, opaque-2, and floury-2 corn endosperms by a modified Osborne procedure. *Journal of Agricultural and Food Chemistry*, *19*(6), pp.1144-1150.

Southgate, E.M., Davey, M.R., Power, J.B. and Marchant, R., 1995. Factors affecting the genetic engineering of plants by microprojectile bombardment. *Biotechnology Advances*, *13*(4), pp.631-651.

Sowokinos, J.R., 2001. Biochemical and molecular control of cold-induced sweetening in potatoes. *American Journal of Potato Research*, *78*(3), pp.221-236.

Sparks, C.A. and Jones, H.D., 2009. Biolistics transformation of wheat. In *Transgenic Wheat, Barley and Oats* (pp. 71-92). *Humana Press*.

Stadler, R.H., Blank, I., Varga, N., Robert, F., Hau, J., Guy, P.A., Robert, M.C. and Riediker, S., 2002. Acrylamide from Maillard reaction products. *Nature*, *419*(6906), pp.449-450.

Steinert, J., Schiml, S., Fauser, F. and Puchta, H., 2015. Highly efficient heritable plant genome engineering using Cas9 orthologues from *Streptococcus thermophilus* and *Staphylococcus aureus*. *The Plant Journal*, *84*(6), pp.1295-1305.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C. and Doudna, J.A., 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, *507*(7490), pp.62-67.

Sumner, S.C., Fennell, T.R., Moore, T.A., Chanas, B., Gonzalez, F. and Ghanayem, B.I., 1999. Role of cytochrome P450 2E1 in the metabolism of acrylamide and acrylonitrile in mice. *Chemical Research in Toxicology*, *12*(11), pp.1110-1116.

Sumner, S.C., Williams, C.C., Snyder, R.W., Krol, W.L., Asgharian, B. and Fennell, T.R., 2003. Acrylamide: a comparison of metabolism and hemoglobin adducts in rodents following dermal, intraperitoneal, oral, or inhalation exposure. *Toxicological Sciences*, *75*(2), pp.260-270.

Sun, H., Wu, S., Zhang, G., Jiao, C., Guo, S., Ren, Y., Zhang, J., Zhang, H., Gong, G., Jia, Z. and Zhang, F., 2017. Karyotype stability and unbiased fractionation in the paleo-allotetraploid Cucurbita genomes. *Molecular Plant*, *10*(10), pp.1293-1306.

Svitashev, S., Schwartz, C., Lenderts, B., Young, J.K. and Cigan, A.M., 2016. Genome editing in maize directed by CRISPR–Cas9 ribonucleoprotein complexes. *Nature Communications*, 7(1), pp.1-7.

Svitashev, S., Young, J.K., Schwartz, C., Gao, H., Falco, S.C. and Cigan, A.M., 2015. Targeted mutagenesis, precise gene editing, and site-specific gene insertion in maize using Cas9 and guide RNA. *Plant Physiology*, *169*(2), pp.931-945.

Symington, L.S. and Gautier, J., 2011. Double-strand break end resection and repair pathway choice. *Annual Review of Genetics*, *45*, pp.247-271.

Ta, T.C., Joy, K.W. and Ireland, R.J., 1985. Role of asparagine in the photorespiratory nitrogen metabolism of pea leaves. *Plant Physiology*, *78*(2), pp.334-337.

Taeymans, D., Wood, J., Ashby, P., Blank, I., Studer, A., Stadler, R.H., Gonde, P., Eijck, P., Lalljie, S.A.M., Lingnert, H. and Lindblom, M., 2004. A review of acrylamide: an industry perspective on research, analysis, formation, and control. *Critical Reviews in Food Science and Nutrition*, *44*(5), pp.323-347.

Tareke, E., Rydberg, P., Karlsson, P., Eriksson, S. and Törnqvist, M., 2000. Acrylamide: a cooking carcinogen?. *Chemical Research in Toxicology*, *13*(6), pp.517-522.

Tareke, E., Rydberg, P., Karlsson, P., Eriksson, S. and Törnqvist, M., 2002. Analysis of acrylamide, a carcinogen formed in heated foodstuffs. *Journal of Agricultural and Food Chemistry*, *50*(17), pp.4998-5006.

Tate, P.H. and Bird, A.P., 1993. Effects of DNA methylation on DNA-binding proteins and gene expression. *Current Opinion in Genetics & Development*, *3*(2), pp.226-231.

Taylor, N.J. and Fauquet, C.M., 2002. Microparticle bombardment as a tool in plant science and agricultural biotechnology. *DNA and Cell Biology*, *21*(12), pp.963-977.

Thyme, S.B., Akhmetova, L., Montague, T.G., Valen, E. and Schier, A.F., 2016. Internal guide RNA interactions interfere with Cas9-mediated cleavage. *Nature Communications*, *7*(1), pp.1-7.

Todd, J., Screen, S., Crowley, J., Peng, J., Andersen, S., Brown, T., Qi, Q., Fabbri, B. and Duff, S.M., 2008. Identification and characterization of four distinct asparagine synthetase (*AsnS*) genes in maize (*Zea mays* L.). *Plant Science*, *175*(6), pp.799-808.

Travella, S., Ross, S.M., Harden, J., Everett, C., Snape, J.W. and Harwood, W.A., 2005. A comparison of transgenic barley lines produced by particle bombardment and Agrobacterium-mediated techniques. *Plant Cell Reports*, *23*(12), pp.780-789.

Tsai, F.Y. and Coruzzi, G.M., 1990. Dark-induced and organ-specific expression of two asparagine synthetase genes in *Pisum sativum*. *The EMBO Journal*, *9*(2), pp.323-332.

Tsai, H., Missirian, V., Ngo, K.J., Tran, R.K., Chan, S.R., Sundaresan, V. and Comai, L., 2013. Production of a high-efficiency TILLING population through polyploidization. *Plant Physiology*, *161*(4), pp.1604-1614.

Tsai, S.Q., Wyvekens, N., Khayter, C., Foden, J.A., Thapar, V., Reyon, D., Goodwin, M.J., Aryee, M.J. and Joung, J.K., 2014. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nature Biotechnology*, *32*(6), pp.569-576.

Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P. and Aryee, M.J., 2015. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature Biotechnology*, *33*(2), p.187.

Twell, D., Klein, T.M., Fromm, M.E. and McCormick, S., 1989. Transient expression of chimeric genes delivered into pollen by microprojectile bombardment. *Plant Physiology*, *91*(4), pp.1270-1274.

Tyl, R.W. and Friedman, M.A., 2003. Effects of acrylamide on rodent reproductive performance. *Reproductive Toxicology*, *17*(1), pp.1-13.

Ueta, R., Abe, C., Watanabe, T., Sugano, S.S., Ishihara, R., Ezura, H., Osakabe, Y. and Osakabe, K., 2017. Rapid breeding of parthenocarpic tomato plants using CRISPR/Cas9. *Scientific Reports*, *7*(1), pp.1-8.

Untergasser, A. et al., 2007. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Research, *35*, pp.71-74.

Upadhyay, S.K., Kumar, J., Alok, A. and Tuli, R., 2013. RNA-guided genome editing for target gene mutations in wheat. *G3: Genes, Genomes, Genetics*, *3*(12), pp.2233-2238.

Urquhart, A.A. and Joy, K.W., 1981. Use of phloem exudate technique in the study of amino acid transport in pea plants. *Plant Physiology*, *68*(3), pp.750-754.

Urquhart, A.A. and Joy, K.W., 1982. Transport, metabolism, and redistribution of xylem-borne amino acids in developing pea shoots. *Plant Physiology*, *69*(5), pp.1226-1232.

USDA-APHIS, 2013. J.R. Simplot Company Petition (13-022-01p) for Determination of Non-regulated Status of Low Acrylamide Potential and Reduced Black Spot Bruise Potato Events F10, F37, E12, E24, J3, J78, G11, H37, and H50: Plant Pest Risk Assessment. USDA-APHIS, Washington DC.

USDA-APHIS, 2014. J.R. Simplot Company Petition (14-093-01p) for Determination of Nonregulated Status for Innate™ Potatoes with Late Blight Resistance, Low Acrylamide Potential, Reduced Black Spot and Lowered Reducing Sugars: Russet Burbank Event W8. USDA-APHIS, Washington DC.

Vain, P., McMullen, M.D. and Finer, J.J., 1993. Osmotic treatment enhances particle bombardment-mediated transient and stable transformation of maize. *Plant Cell Reports*, *12*(2), pp.84-88.

van Oeveren, J., de Ruiter, M., Jesse, T., van der Poel, H., Tang, J., Yalcin, F., Janssen, A., Volpin, H., Stormo, K.E., Bogden, R. and van Eijk, M.J., 2011. Sequence-based physical mapping of complex genomes by whole genome profiling. *Genome Research*, *21*(4), pp.618-625.

VanBuren, R., Bryant, D., Bushakra, J.M., Vining, K.J., Edger, P.P., Rowley, E.R., Priest, H.D., Michael, T.P., Lyons, E., Filichkin, S.A. and Dossett, M., 2016. The genome of black raspberry (*Rubus occidentalis*). *The Plant Journal*, *87*(6), pp.535-547.

Vasil, V. and Vasil, I.K., 1980. Isolation and culture of cereal protoplasts. *Theoretical and Applied Genetics*, *56*(3), pp.97-99.

Vasil, V., Castillo, A.M., Fromm, M.E. and Vasil, I.K., 1992. Herbicide resistant fertile transgenic wheat plants obtained by microprojectile bombardment of regenerable embryogenic callus. *Bio/technology*, *10*(6), pp.667-674.

Von Tungeln, L.S., Doerge, D.R., Gamboa da Costa, G., Matilde Marques, M., Witt, W.M., Koturbash, I., Pogribny, I.P. and Beland, F.A., 2012. Tumorigenicity of acrylamide and its metabolite glycidamide in the neonatal mouse bioassay. *International Journal of Cancer*, *131*(9), pp.2008-2015.

Vouillot, L., Thélie, A. and Pollet, N., 2015. Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3: Genes, Genomes, Genetics*, *5*(3), pp.407-415.

Voytas, D.F., 2013. Plant genome engineering with sequence-specific nucleases. *Annual Review of Plant Biology*, *64*.

Vu, T.V., Sivankalyani, V., Kim, E.J., Doan, D.T.H., Tran, M.T., Kim, J., Sung, Y.W., Park, M., Kang, Y.J. and Kim, J.Y., 2020. Highly efficient homology-directed repair using CRISPR/Cpf1-geminiviral replicon in tomato. *Plant Biotechnology Journal*.

Wang, F., Wang, C., Liu, P., Lei, C., Hao, W., Gao, Y., Liu, Y.G. and Zhao, K., 2016. Enhanced rice blast resistance by CRISPR/Cas9-targeted mutagenesis of the ERF transcription factor gene *OsERF922*. *PloS One*, *11*(4), p.e0154027.

Wang, H., Liu, D., Sun, J. and Zhang, A., 2005. Asparagine synthetase gene *TaASN1* from wheat is up-regulated by salt stress, osmotic stress and ABA. *Journal of Plant Physiology*, *162*(1), pp.81-89.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F. and Jaenisch, R., 2013. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*, *153*(4), pp.910-918.

Wang, M., Lu, Y., Botella, J.R., Mao, Y., Hua, K. and Zhu, J.K., 2017. Gene targeting by homology-directed repair in rice using a geminivirus-based CRISPR/Cas9 system. *Molecular Plant*, *10*(7), pp.1007-1010.

Wang, W., Pan, Q., He, F., Akhunova, A., Chao, S., Trick, H. and Akhunov, E., 2018. Transgenerational CRISPR-Cas9 activity facilitates multiplex gene editing in allopolyploid wheat. *The CRISPR Journal*, *1*(1), pp.65-74.

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C. and Qiu, J.L., 2014. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nature Biotechnology*, *32*(9), p.947.

Wang, Y.C., Klein, T.M., Fromm, M., Cao, J., Sanford, J.C. and Wu, R., 1988. Transient expression of foreign genes in rice, wheat and soybean cells following particle bombardment. *Plant Molecular Biology*, *11*(4), pp.433-439.

Wang, Z.P., Xing, H.L., Dong, L., Zhang, H.Y., Han, C.Y., Wang, X.C. and Chen, Q.J., 2015. Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in Arabidopsis in a single generation. *Genome Biology*, *16*(1), p.144.

Warth, B., Parich, A., Bueschl, C., Schoefbeck, D., Neumann, N.K.N., Kluger, B., Schuster, K., Krska, R., Adam, G., Lemmens, M. and Schuhmacher, R., 2015. GC–MS based targeted metabolic profiling identifies changes in the wheat metabolome following deoxynivalenol treatment. *Metabolomics*, *11*(3), pp.722-738.

Watanabe, T., Shimbo, S., Moon, C.S., Zhang, Z.W. and Ikeda, M., 1996. Cadmium contents in rice samples from various areas in the world. *Science of the Total Environment*, *184*(3), pp.191-196.

Waterhouse, P.M. and Helliwell, C.A., 2003. Exploring plant genomes by RNA-induced gene silencing. *Nature Reviews Genetics*, *4*(1), pp.29-38.

Weißhaar, R., 2004. Acrylamide in heated potato products–analytics and formation routes. *European Journal of Lipid Science and Technology*, *106*(11), pp.786-792.

Wiberley-Bradford, A.E. and Bethke, P.C., 2018. Suppression of the vacuolar invertase gene delays senescent sweetening in chipping potatoes. *Journal of the Science of Food and Agriculture*, *98*(1), pp.354-360.

Wilkinson, P.A., Winfield, M.O., Barker, G.L., Tyrrell, S., Bian, X., Allen, A.M., Burridge, A., Coghill, J.A., Waterfall, C., Caccamo, M. and Davey, R.P., 2016. CerealsDB 3.0: expansion of resources and data integration. *BMC Bioinformatics*, *17*(1), p.256.

Winkler, U. and Schön, W.J., 1980. Amino acid composition of the kernel proteins in barley resulting from nitrogen fertilization at different stages of development. *Journal of Agronomy and Crop Science*, *149*, pp.503-512.

Wong, H.K., Chan, H.K., Coruzzi, G.M. and Lam, H.M., 2004. Correlation of *ASN2* gene expression with ammonium metabolism in Arabidopsis. *Plant Physiology*, *134*(1), pp.332-338.

Wong, G.K.S., Wang, J., Tao, L., Tan, J., Zhang, J., Passey, D.A. and Yu, J., 2002. Compositional gradients in Gramineae genes. *Genome Research*, *12*(6), pp.851-856.

Woo, J.W., Kim, J., Kwon, S.I., Corvalán, C., Cho, S.W., Kim, H., Kim, S.G., Kim, S.T., Choe, S. and Kim, J.S., 2015. DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nature Biotechnology*, *33*(11), pp.1162-1164.

Wright, D.A., Townsend, J.A., Winfrey Jr, R.J., Irwin, P.A., Rajagopal, J., Lonosky, P.M., Hall, B.D., Jondle, M.D. and Voytas, D.F., 2005. High-frequency homologous recombination in plants mediated by zinc-finger nucleases. *The Plant Journal*, *44*(4), pp.693-705.

Wu, L., Nandi, S., Chen, L., Rodriguez, R.L. and Huang, N., 2002. Expression and inheritance of nine transgenes in rice. *Transgenic Research*, *11*(5), pp.533-541.

Wu, C., Li, X., Yuan, W., Chen, G., Kilian, A., Li, J., Xu, C., Li, X., Zhou, D.X., Wang, S. and Zhang, Q., 2003. Development of enhancer trap lines for functional analysis of the rice genome. *The Plant Journal*, *35*(3), pp.418-427.

Wu, X., Kriz, A.J. and Sharp, P.A., 2014a. Target specificity of the CRISPR-Cas9 system. *Quantitative Biology*, *2*(2), pp.59-70.

Wu, X., Scott, D.A., Kriz, A.J., Chiu, A.C., Hsu, P.D., Dadon, D.B., Cheng, A.W., Trevino, A.E., Konermann, S., Chen, S. and Jaenisch, R., 2014b. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature Biotechnology*, *32*(7), pp.670-676.

Xiao, A., Cheng, Z., Kong, L., Zhu, Z., Lin, S., Gao, G. and Zhang, B., 2014. CasOT: a genome-wide Cas9/gRNA off-target searching tool. *Bioinformatics*, *30*(8), pp.1180-1182.

Xie, K. and Yang, Y., 2013. RNA-guided genome editing in plants using a CRISPR–Cas system. *Molecular Plant*, *6*(6), pp.1975-1983.

Xie, K., Minkenberg, B. and Yang, Y., 2015. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proceedings of the National Academy of Sciences*, *112*(11), pp.3570-3575.

Xing, H.L., Dong, L., Wang, Z.P., Zhang, H.Y., Han, C.Y., Liu, B., Wang, X.C. and Chen, Q.J., 2014. A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biology*, *14*(1), p.327.

Xu, G., Fan, X. and Miller, A.J., 2012. Plant nitrogen assimilation and use efficiency. *Annual Review of Plant Biology*, *63*, pp.153-182.

Xu, H., Curtis, T.Y., Powers, S.J., Raffan, S., Gao, R., Huang, J., Heiner, M., Gilbert, D.R. and Halford, N.G., 2018. Genomic, biochemical, and modeling analyses of asparagine synthetases from wheat. *Frontiers in Plant Science*, *8*, p.2237.

Xu, K., Ren, C., Liu, Z., Zhang, T., Zhang, T., Li, D., Wang, L., Yan, Q., Guo, L., Shen, J. and Zhang, Z., 2015. Efficient genome engineering in eukaryotes using Cas9 from *Streptococcus thermophilus*. *Cellular and Molecular Life Sciences*, *72*(2), pp.383-399.

Yan, L., Wei, S., Wu, Y., Hu, R., Li, H., Yang, W. and Xie, Q., 2015. High-efficiency genome editing in Arabidopsis using YAO promoter-driven CRISPR/Cas9 system. *Molecular Plant*, *8*(12), pp.1820-1823.

Yu, J., Hu, S., Wang, J., Wong, G.K.S., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X. and Cao, M., 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science*, *296*(5565), pp.79-92.

Zambryski, P., Holsters, M., Kruger, K., Depicker, A., Schell, J., Van Montagu, M. and Goodman, H.M., 1980. Tumor DNA structure in plant cells transformed by *A. tumefaciens*. *Science*, *209*(4463), pp.1385-1391.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., Van Der Oost, J., Regev, A. and Koonin, E.V., 2015. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, *163*(3), pp.759-771.

Zhan, J., Thakare, D., Ma, C., Lloyd, A., Nixon, N.M., Arakaki, A.M., Burnett, W.J., Logan, K.O., Wang, D., Wang, X. and Drews, G.N., 2015. RNA sequencing of laser-capture microdissected compartments of the maize kernel identifies regulatory modules associated with endosperm cell differentiation. *The Plant Cell*, *27*(3), pp.513-531.

Zhang, F., LeBlanc, C., Irish, V.F. and Jacob, Y., 2017. Rapid and efficient CRISPR/Cas9 gene editing in Citrus using the YAO promoter. *Plant Cell Reports*, *36*(12), pp.1883-1887.

Zhang, F., Maeder, M.L., Unger-Wallace, E., Hoshaw, J.P., Reyon, D., Christian, M., Li, X., Pierick, C.J., Dobbs, D., Peterson, T. and Joung, J.K., 2010. High frequency targeted mutagenesis in *Arabidopsis thaliana* using zinc finger nucleases. *Proceedings of the National Academy of Sciences*, *107*(26), pp.12028-12033.

Zhang, H., Zhang, J., Wei, P., Zhang, B., Gou, F., Feng, Z., Mao, Y., Yang, L., Zhang, H., Xu, N. and Zhu, J.K., 2014a. The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnology Journal*, *12*(6), pp.797-807.

Zhang, Q., Zhang, X., Wang, S., Tan, C., Zhou, G. and Li, C., 2016a. Involvement of alternative splicing in barley seed germination. *PLoS One*, 11(3), p.e0152824.

Zhang, Y., Dickinson, J.R., Paul, M.J. and Halford, N.G., 2003. Molecular cloning of an Arabidopsis homologue of GCN2, a protein kinase involved in co-ordinated response to amino acid starvation. *Planta*, *217*(4), pp.668-675.

Zhang, Y., Heidrich, N., Ampattu, B.J., Gunderson, C.W., Seifert, H.S., Schoen, C., Vogel, J. and Sontheimer, E.J., 2013. Processing-independent CRISPR RNAs limit natural transformation in Neisseria meningitidis. *Molecular Cell*, *50*(4), pp.488-503.

Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., Qiu, J.L. and Gao, C., 2016b. Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nature Communications*, *7*(1), pp.1-8.

Zhang, Y., Wang, Y., Kanyuka, K., Parry, M.A., Powers, S.J. and Halford, N.G., 2008. GCN2-dependent phosphorylation of eukaryotic translation initiation factor-2α in Arabidopsis. *Journal of Experimental Botany*, *59*(11), pp.3131-3141.

Zhang, Y.C., Liao, J.Y., Li, Z.Y., Yu, Y., Zhang, J.P., Li, Q.F., Qu, L.H., Shu, W.S. and Chen, Y.Q., 2014b. Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biology*, *15*(12), p.512.

Zhang, Z., Hua, L., Gupta, A., Tricoli, D., Edwards, K.J., Yang, B. and Li, W., 2019. Development of an Agrobacterium-delivered CRISPR/Cas9 system for wheat genome editing. *Plant Biotechnology Journal*, *17*(8), pp.1623-1635.

Zhao, F.J., Hawkesford, M.J. and McGrath, S.P., 1999. Sulphur assimilation and effects on yield and quality of wheat. *Journal of Cereal Science*, *30*(1), pp.1-17.

Zhao, Y., Zhang, C., Liu, W., Gao, W., Liu, C., Song, G., Li, W.X., Mao, L., Chen, B., Xu, Y. and Li, X., 2016. An alternative strategy for targeted gene replacement in plants using a dual-sgRNA/Cas9 design. *Scientific Reports*, *6*(1), pp.1-11.

Zheng, G.X., Lau, B.T., Schnall-Levin, M., Jarosz, M., Bell, J.M., Hindson, C.M., Kyriazopoulou-Panagiotopoulou, S., Masquelier, D.A., Merrill, L., Terry, J.M. and Mudivarti, P.A., 2016a. Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nature Biotechnology*, *34*(3), pp.303-311.

Zheng, X., Yang, S., Zhang, D., Zhong, Z., Tang, X., Deng, K., Zhou, J., Qi, Y. and Zhang, Y., 2016b. Effective screen of CRISPR/Cas9-induced mutants in rice by single-strand conformation polymorphism. *Plant Cell Reports*, *35*(7), pp.1545-1554.

Zhivagui, M., Ng, A.W., Ardin, M., Churchwell, M.I., Pandey, M., Renard, C., Villar, S., Cahais, V., Robitaille, A., Bouaoun, L. and Heguy, A., 2019. Experimental and pan-cancer genome analyses reveal widespread contribution of acrylamide exposure to carcinogenesis in humans. *Genome Research*, *29*(4), pp.521-531.

Zhou, H., Liu, B., Weeks, D.P., Spalding, M.H. and Yang, B., 2014. Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. *Nucleic Acids Research*, *42*(17), pp.10903-10914.

Zhu, X., Richael, C., Chamberlain, P., Busse, J.S., Bussan, A.J., Jiang, J. and Bethke, P.C., 2014. Vacuolar invertase gene silencing in potato (*Solanum tuberosum* L.) improves processing quality by decreasing the frequency of sugar-end defects. *PloS One*, *9*(4), p.e93381.

Žilić, S., Dodig, D., Basić, Z., Vančetović, J., Titan, P., Đurić, N. and Tolimir, N., 2017. Free asparagine and sugars profile of cereal species: the potential of cereals for acrylamide formation in foods. *Food Additives & Contaminants: Part A*, *34*(5), pp.705-713.

Zimin, A.V., Puiu, D., Hall, R., Kingan, S., Clavijo, B.J. and Salzberg, S.L., 2017. The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum*. *Gigascience*, *6*(11), p.gix097.

Zyzak, D.V., Sanders, R.A., Stojanovic, M., Tallmadge, D.H., Eberhart, B.L., Ewald, D.K., Gruber, D.C., Morsch, T.R., Strothers, M.A., Rizzi, G.P. and Villagran, M.D., 2003. Acrylamide formation mechanism in heated foods. *Journal of Agricultural and Food Chemistry*, *51*(16), pp.4782-4787.